THE ACTIVATION MECHANISM OF RHODOPSIN EXPLORED BY MULTISCALE METHODS

by

Basak Isin

BS, Chemical Engineering, Bogazici University, 1998

MS, Chemical Engineering, Bogazici University, 2000

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Basak Isin

It was defended on

[08 07, 2007]

and approved by

Signature	Printed Name	Date Signed
_____ (Committee member)	Dr. Michael Cascio, Assistant Professor	_____
_____ (Committee member)	Dr. Hagai Meirovitch, Professor	_____
_____ (Committee member)	Dr. Judith Klein-Seetharaman, Assistant Professor	_____
_____ (Committee member)	Dr. Thomas E. Smithgall, Professor	_____
_____ (Dissertation Director)	Dr. Ivet Bahar, Professor	_____

THE ACTIVATION MECHANISM OF RHODOPSIN EXPLORED BY MULTISCALE METHODS

Basak Isin, M.S.

University of Pittsburgh, 2007

Rhodopsin is the best characterized member of the large, pharmaceutically important, family of G-protein-coupled receptors (GPCRs), and serves as a prototype for understanding GPCR activation. In this thesis, we aim at understanding the activation mechanism of rhodopsin. To this aim, we first performed an in-depth analysis of the conformational motions of rhodopsin predicted by two elastic network models, Gaussian Network Model (GNM) and Anisotropic Network Model (ANM). We compared these motions with the extensive amount of experimental data, and developed a model for rhodopsin activation. We tested the model with Meta II fluorescence decay rates measured to characterize the deactivation of rhodopsin mutants. We find that our results correctly predict 93% of the experimentally observed effects in 119 rhodopsin mutants for which the decay rates and misfolding data were measured, including a systematic analysis of Cys->Ser replacements. Next, in order to incorporate atomic details and the effects of membrane and water molecules into our model, we developed a new protocol named ANM-restrained molecular dynamics (MD). In this protocol, we used multiple ANM modes as restraints to guide MD simulations. By using this protocol, we were able to sample biologically relevant, large scale motions of the protein that are otherwise not accessible to the conventional timescales MD simulations. Furthermore, we explored the evolution of the multiple ANM global modes with realistic deformations favored by a detailed atomic force field in the presence of the explicit environment. Remarkably, with this method, we identify a highly hinge site, which does not change with several rounds of applying normal modes as restraints. This hinge site includes residues that are directly affected by the isomerization of retinal, as well as

those stabilizing the resulting all-*trans* conformation of the chromophore. The CP ends of the helices H3, H4, H5, and H6 and the connecting loops are found to enjoy an enhanced mobility facilitated by this hinge site. Several new interactions are observed to contribute to the mechanism of signal propagation from the retinal binding pocket to the G-protein binding sites in the CP domain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEGMENTS

I would like to express my appreciation and gratitude to my mentor Dr. Ivet Bahar for intriguing my interest in molecular biophysics and opening me a new area in life. Her dedication to science and her hard work will continue to be an inspiration to me. She has also provided me freedom that allowed me to take my Ph.D. project into a direction that seemed necessary to me.

I have had the fortunate opportunity to work with Dr. Judith Klein-Seetheraman whose extensive knowledge of biology and experimental techniques I benefited immensely. I am also very impressed by her unique, outspoken and humanitarian nature which allowed me to stay optimistic throughout all the hardships.

I extend my thanks to the members of my thesis committee members, Dr. Hagai Meirowitch, Dr. Mike Cascio and Dr. Tom Smithgall for their constructive suggestions. I thank Dr. AJ Rader for his contributions to the manuscript that is relevant to Chapter 2 of this thesis. I would like to thank Dr. Steve Phillips for his counseling and assistance entering the interdisciplinary biomedical graduate program. I thank to Dr. Rajan Munshi who provided a good basis for the learning process during my transition from being a chemical engineer to a biologist.

I would like to thank my most intellectual friend and colleague Dr. Chakra Chennubhotla for his help and for useful scientific discussions that we had throughout my PhD as well as his critical reading of this thesis.

I would like to thank Dr. Klaus Schulten and Dr. Emad Tajkhorshid for accepting me as a visiting student to Theoretical and Computational Biophysics Group at University of Illinois at Urbana-Champaign. It has been both an honor and a privilege for me to be with such a dedicated and talented team. I have gained experience in powerful contemporary techniques in computational biology during my visit.

I thank the members of my family: my mother Zubeyde Isin, my father Okan Isin, my sister Burcu Isin, my aunt Dr. Cigdem Gulal Diler and her husband Dr. Yunus Diler, my grandmothers Guzin Isin and Cemile Gulal, my grandmother Osman Gulal, and my uncle Tamer Gulal. With their unconditional support and love only an outstanding family could provide, I have been able to keep focused on my goals in life. My family has been a pillar of strength for me from the day I left Turkey and has always been there for me since I embarked on my graduate school journey. I especially would like thank my grandfather Osman Gulal whose discipline and hard work taught me to strive to achieve my goals.

I feel very lucky to have friends like Amie Eisfeld-Fenney, Curtis Fenney and J. Trevor Lester. Thanks to their generous spirit and guidance, I was able to adjust easily and enjoy the life in US and most importantly learn how to become an independent individual. I would like give my warmest thanks to my dear Umut Arslan who allowed me to stay cheerful and strong even at difficult times with his unwavering support. I also thank my friends who became my second family in Pittsburgh: Yesim Basartan, Zeynep Sevener, Gorkem Saka, Aysegul Siranli, Cassandra Storm, Tuba Pinar Yildirim, Ahmet Bakan, Kasey Eidson, Amy Gardiner, Isil Dilek, Marie Acquilano, Ceyda Acilan, Baris Taskin, Oznur Tastan, Nina Korzeniewski, Rick Blakesly, John Shaffer, Benhur Aysin, Elif Okut-Aysin, Alpay Temiz, Kadir Diri, and Mehmet Can Demirci.

1.0 INTRODUCTION

1.1 PROTEIN DYNAMICS

Protein sequences fold into unique three dimensional structures to perform their biological function (Finkelstein and Ptitsyn, 2002). Structural information on proteins is rapidly accumulating due to efficient techniques such as X-ray crystallography and nuclear magnetic resonance (NMR). While the structure provides insights into function, it only yields static information. Detailed information on the dynamics of a protein is necessary for a complete understanding of its function. However, it is not a trivial problem to extract functionally relevant motions of biomolecular systems. The time scales involved in the dynamics of different properties range from femtoseconds to seconds or even longer (Table 1.1). When undertaking a biomolecular modeling study of a particular system of interest, the level of modeling, that is, the spatial resolution, time scale, and the degrees of freedom of interest, must be considered. Simulation techniques can help to obtain dynamic information that cannot be provided by experimental techniques in a straightforward manner. Computational biology tools and biophysical theories have been useful in investigating the dynamics of biomolecular systems when they are used in conjunction with experimental data. Molecular dynamics (MD), Normal mode analysis (NMA) are the most widely used computational techniques (Becker et al., 2001;Leach, 2001;Schlick, 2002;Cui and Bahar, 2006). Here, we will briefly introduce these

methods and details are presented in the method sections of the relevant chapters (Chapter 2, 3 and 4).

Table 1.1 Timescales and amplitudes for characteristic motions of proteins

Types of motions	Functionality examples	Amplitude of motions (Å)	Time scales (sec)
Local motions Atomic fluctuations Side chain motion	Ligand docking flexibility Temporal diffusion pathways	1-3	$< 10^{-9}$
Medium-scale motions Loop motion Terminal-arm motion	Active site conformation adaptation Binding specificity	3-8	10^{-9}-10^{-8}
Large-scale motions Domain movements Subunit rearrangements	Hinge-bending motion Allosteric transitions	8-20	10^{-8}-10^{-5}
Global motions Folding/ unfolding Helix-coil transition Subunit association	Hormone activation Protein functionality	>20	$< 10^{-5}$

1.1.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are routinely used today to study the time-dependent behavior of molecular systems. They investigate motions of a system of discrete particles under the influence of internal and external forces. The interactions of the particles are empirically described by a potential energy function from which the forces that act on each particle are derived. With knowledge of these forces, it is possible to calculate the dynamic behavior of the system using classical Newton's equations of motion for each atom in the system (Becker et al., 2001). These fine grained simulations provide atomic-level detail with high temporal resolution for both harmonic and anharmonic motions. For biomolecular systems, a time step of a few femtoseconds is used and a typical simulation consists of millions of steps. The rapid growth in computer power and new developments in parallel MD codes designed for high-performance simulations enable us to simulate large biomolecular systems at hundreds of nanoseconds to

microseconds scales at best (Phillips et al., 2005). However, slow conformational transitions which are important for the functional motions of proteins and folding processes are in the order of micro to milliseconds and these timescales are too long for MD simulations (Table 1.1). Furthermore, the inefficient conformational sampling due to being trapped in a local minimum for a long time is another problem for MD simulations. Hence, standard MD is not a particularly efficient method for sampling large conformational changes spanning periods of time longer than microseconds especially for large macromolecules.

Various methods have been developed to accelerate MD simulations and increase their sampling efficiency. Steered Molecular Dynamics (SMD) method has been used successfully to observe large conformational changes by applying external forces (Isralewitz et al., 2001). A Similar approach, Targeted Moleculer Dynamics (TMD), applies time-dependent, purely geometrical constraints to induce conformational changes for achieving a known target structure at ordinary temperature. Additionally, guiding MD simulations by collective coordinates has proven to be useful for efficient sampling and expansion of the accessible conformational space (Kitao and Go, 1999). Berendsen and coworkers developed an essential dynamics simulation method in which essential dynamics analysis (EDA) or principle component analysis (PCA) of a predefined group of conformations obtained from MD simulation trajectories are used to generate collective modes (Amadei et al., 1993). These collective modes are used as constraints in MD simulations to sample the conformational space efficiently (Amadei et al., 1993;Amadei et al., 1996;de Groot et al., 1996). Collective coordinates have been combined with ensemble sampling by Abheser and Nilges. Their method uses an additional biasing potential defined along the collective modes used as restraints in a set of independent MD trajectories(Abseher and Nilges, 2000). Another technique, Amplified Collective Motion couples low frequency motions

obtained from the coarse grained anisotropic network model, to a higher temperature by using a weak coupling method (Zhang et al., 2003).

1.1.2 Normal mode analysis and Elastic Network Models

Normal Mode Analysis (NMA) is an efficient method for identifying collective motions of biomolecules at long timescales and for enabling to visualize biologically interesting modes. While describing all bonded and nonbonded interactions of a molecule by semi-empirical potential functions as in MD (Figure 1.1A), NMA is based on a harmonic approximation and assumes that the conformational energy surface can be characterized by a single energy minimum. The quality of the modes obtained is dependent on reaching a true minimum on the potential energy surface. Hence, a stringent energy minimization of the potential energy function is required before performing NMA. Afterwards, the Hessian matrix is generated by the second derivatives of the potential energy function with respect to atomic positions. and diagonalized to obtain the collective modes. The low-frequency modes often correlate well with experimentally observed conformational changes associated with function. However, energy minimization and diagonalization of the Hessian Matrix makes NMA a computationally expensive method for large biomolecules (Berendsen and Hayward, 2000;Becker et al., 2001).

Instead of using a complex potential function, Tirion showed the low frequency modes of motion can be satisfactorily obtained by an NMA with a single parameter harmonic potential (Figure 1.1B) (Tirion, 1996). Furthermore, this method eliminates the energy minimization step. Bahar and co-workers took the simplification further by representing each residue with a single node (Figure 1.1C) and developed two Elastic Network Analyses (ENM), Gaussian Network

Model (GNM) (Bahar et al., 1997) and Anisotropic Network Model (ANM) (Doruker et al., 2000; Atilgan et al., 2001). The basic ingredient in these models is the topology of inter-residue contacts in the native structure, which turns out to be a major factor defining equilibrium dynamics. The GNM and ANM have been shown to accurately reproduce experimental residue fluctuations. The interactions between residues in close proximity are represented by harmonic potentials with a uniform spring constant, and the network junctions are usually identified by the C^{α} atoms (Bahar et al., 1997; Hinsen et al., 1998; Doruker et al., 2000; Atilgan et al., 2001). The GNM analysis provides information on the *size* of collective motions. The three dimensional extension, the ANM (Atilgan et al., 2001), provides useful information about the directions and magnitudes of global motions (Keskin et al., 2002;Isin et al., 2002). These models have been widely used in recent years for exploring the biologically relevant, long timescale motions of large structures while avoiding expensive computations (for reviews see Bahar and Rader (2005); Chennubhotla et al., 2005; Ma, 2005).

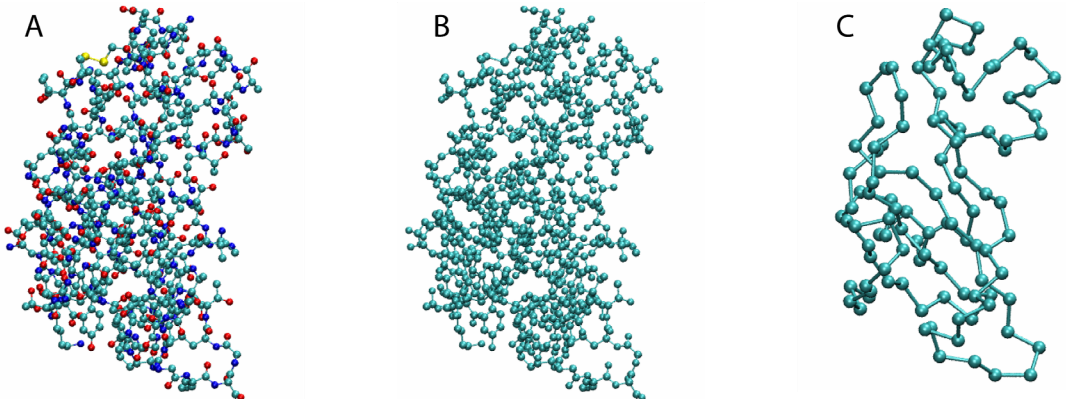

Figure 1.1 Representation of proteins at different resolutions:
A. NMA: Every atom constituting the protein is used in NMA calculations and semiempirical potentials characterizes the covalent and noncovalent interactions between atoms. B. Tirion's Model: The complex potential is replaced by a simple pairwise Hookean potential. All atoms of the protein is included in the calculations. C. ENMs: Each residue in the protein is represented by a single node. Nodes that are sufficiently close are connected by a uniform spring.

1.1.2.1 Global modes

The distributions of residue fluctuations driven by the lowest frequency modes are referred to as 'global' modes. They are shown to be insensitive to the details of the models and energy parameters used in normal mode analyses, provided that the distribution of inter-residue contacts is maintained (Tama and Sanejouand, 2001;Lu and Ma, 2005). Empirical studies of the global modes of a number of different protein structures indicate that *minima* in the global mode coincide with the hinge sites of the molecule, while *maxima* usually correspond to substrate recognition sites. This observation indicates that in the global modes, the hinge sites are fixed in space, whereas substrate recognition sites sample a large conformational space. The mechanical stability of the hinge sites is a requirement for them to serve as a swivel, a pivot, or a shaft, about which the collective motions are actuated, while the high flexibility of the recognition sites facilitates the recognition and optimal binding of substrates. Both, hinge regions as well as substrate recognition sites are in general critical for biological function. For example, the highly conserved shallow pockets that serve as receptor binding sites in influenza virus hemagglutinin A, or the antigen binding hypervariable loops of immunoglobulins, form maxima in the slowest mode shape, whereas linkers or interfacial regions between domains subject to anticorrelated motions form minima (Ming et al., 2002;Chennubhotla et al., 2005;Bahar and Rader, 2005;Ma, 2005).

1.1.2.2 High frequency fluctuations

In contrast to global modes, high frequency modes are sensitive to the detailed interactions at the atomic level. They usually contain white noise contributions that need to be filtered out in order to extract physically meaningful information. Not surprisingly, these modes have been referred

to as ‘uninteresting modes’ (Amadei et al., 1993). They usually drive *isolated* fluctuations, as opposed to the *correlated* ones that underlie the intramolecular communication.

The ENM results differ from those extracted from conventional simulations in that they are devoid of random noise effects. The high frequency modes identified by the ENM are ‘interesting’: they indicate the most strongly constrained sites in the presence of the intricate coupling between *all* residues. The peaks emerging in these mode shapes are usually implicated in folding nuclei, or key tertiary contacts stabilizing the overall fold. As a consequence, they ought to be evolutionarily conserved among different members of a given family (Demirel et al., 1998;Bahar et al., 1998).

1.2 RHODOPSIN AND ITS DYNAMICS

G protein–coupled receptors (GPCRs) are the largest known superfamily of receptors. They are involved in a number of clinically important ligand-receptor processes and perform diverse functions including responses to light, odorant molecules, neurotransmitters, hormones and a variety of other signals. 50-60% of all approved drugs target members of the GPCR family. Hence, GPCRs are pharmacologically important. A better understanding of the functions and dynamics of these receptors and their structure will help in the design of drugs for the treatment of GPCRs-related diseases. Being highly abundant in nature, rhodopsin is the most widely studied and the best-characterized member of GPCR family for which only accurate structural information is available (Palczewski et al., 2000;Okada et al., 2002;Salom et al., 2006). The structure-function studies of rhodopsin provide the fundamental basis for understanding how members of the GPCR family work.

Rhodopsin is the dim light photoreceptor located at the outer segments of rod photoreceptor cells in the retina. It responds to photons by initiating a series of intracellular processes that result in an electrical signal processed by the visual system (Hubbell et al., 2003).

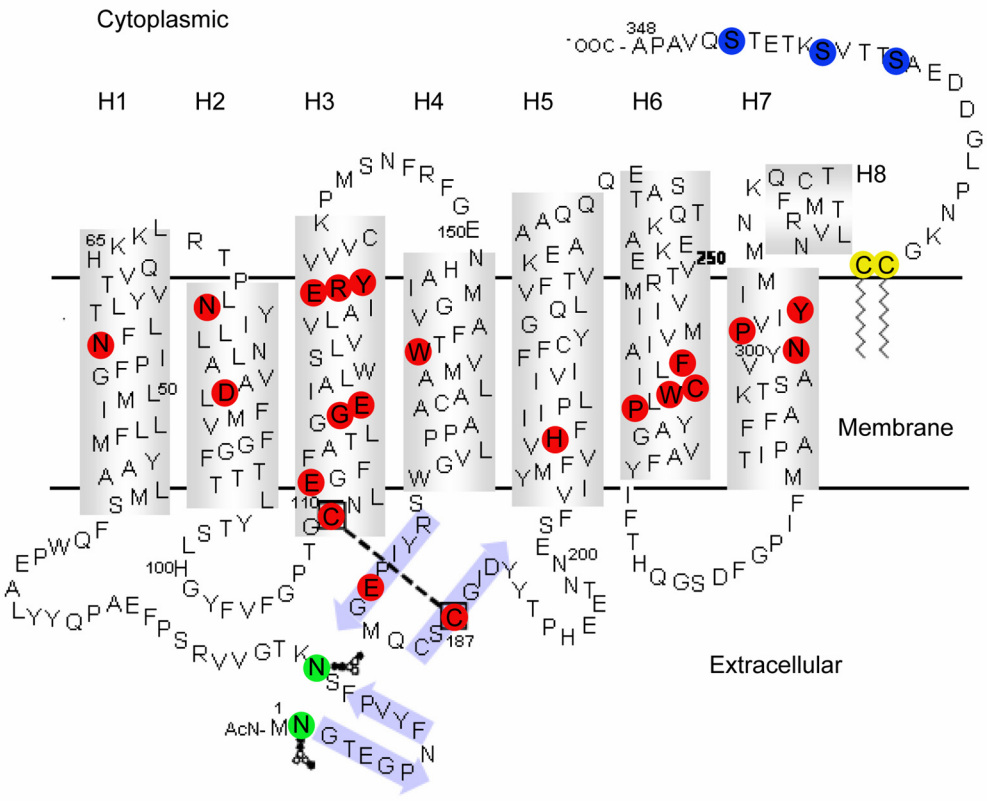

Figure 1.2 Secondary structure diagram of rhodopsin
The seven transmembrane (TM) helices (H1 to H7) and the soluble H8 are shown as grey cylinders. Residues are depicted in single-letter code. The carboxyl-terminus and the CP domain are located at the top, while the amino-terminus and EC are at the bottom. EC loops fold into TM region as β strands and the β strand at the N terminus are shown by arrows. An essential disulfide bond linking Cys110 and Cys187 at the interface between TM and EC domain is indicated by dashed lines. Highly conserved residues among GPCRs including D(E)RY and NPXXY motives at H3 and H8 are colored red. The palmitoylated Cys-322 and Cys-323 (yellow) anchor to the membrane and provide constraints for H8 at the CP domain. Asn2 and Asn15 (green) are the glycosylation sites. Main phosphorylation sites are Ser334, Ser338, and Ser343 at the C terminus (colored blue).

1.2.1 Rhodopsin structure

Like all GPCRs, rhodopsin comprises cytoplasmic (CP), transmembrane (TM), extracellular (EC) domains (Figure1.1 and 1.2A-B) and contains a bundle of seven TM helices (H1-H7) originally observed by cryo-EM (Unger and Schertler, 1995) and confirmed later by the X-ray crystal structures in the inactive or dark state (Figure 1.2) (Palczewski et al., 2000;Okada and Palczewski, 2001;Teller et al., 2001;Okada et al., 2002;Okada et al., 2004). This TM bundle

encloses the chromophore, 11-*cis*-retinal, covalently bound to the ε-amino group of Lys296 at H7 through a Schiff-base linkage. 11-*cis*-retinal (Figure 1.2C) is a derivative of vitamin A_1, consisting of a β-ionone ring (carbon atoms C1-C6), a polyene chain (C7-C15) and several methyl groups (C16-C20). EC2 forms a β-strand that serves as a lid to the chromophore, running almost parallel to the polyene chain of the retinal and stabilized by a highly conserved disulfide bond between Cys110 and Cys187 which was shown to be important for stability of both active and inactive conformations as well as the correct folding of the protein (Palczewski et al., 2000). The CP domain of rhodopsin contains three interhelical loops, CL1-CL3, connecting the respective pairs of helices H1-H2, H3-H4, and H5-H6. Additionally, there is a soluble helix H8, in the CP domain, which runs parallel to the plane of the membrane. The amino acid sequence of rhodopsin molecule and the orientation of the structural elements in the membrane are shown in Figure 1.1. Several conserved and important residues shown to play a role in chromophore binding and receptor activation are also indicated by red circles. Cys322 and Cys323 in the CP domain are attached to the membrane by palmitoylation. Ser334, Ser338, and Ser343 are the main phosphorylation sites. The EC domain consists of the N-terminus and three interhelical loops (EC1-EC3) between TM helices H2-H3, H4-H5 and H6-H7, respectively. At the N-terminus, Met1 is acetylated. Asn2 and Asn15 are the glycosylation sites.

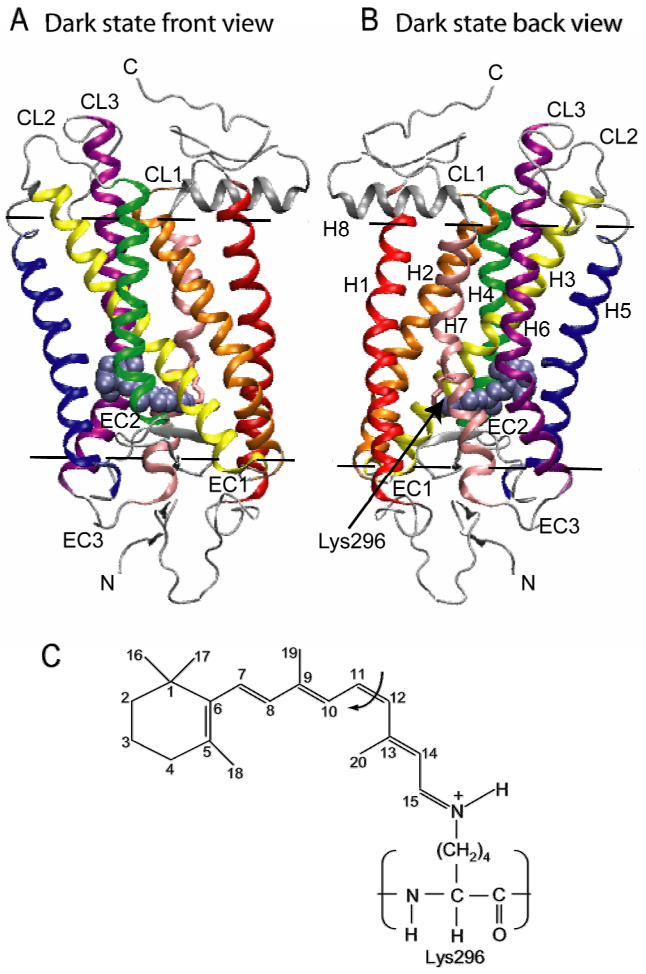

Figure 1.3 Structure of rhodopsin in the dark state.
(A) Front view. (B) Back view of the PDB structure 1U19 deposited by Okada et al (Okada et al., 2004) The structure contains seven TM helices: Helix 1 (H1; residues 33-65), red; Helix 2 (H2; 70-101), orange; Helix3 (H3; residues 105-140), yellow; Helix 4 (H4; residues 149-173), green; Helix 5 (H5; residues 199-226), blue; Helix 6 (H6; residues 243-278), magenta; and Helix 7 (H7; residues 284-310), pink. The non-helical portions and the CP helix 8 (H8) are in gray. 11-*cis*-retinal is shown in light-blue space-filling model. The dashed black lines show the approximate boundaries of the hydrophobic core of the membrane. The CP and EC regions contain the respective loops CL1-CL3 and EC1-EC3. The side chain of Lys296 is displayed in the front and back view. (C) 11-*cis*-retinal and the protonated Schiff base linkage to Lys296. The arrow indicates the bond subject to *cis-trans* isomerization. This image was generated using VMD (Humphrey et al., 1996).

1.2.2 Photointermediates of Rhodopsin

Its chromophore, 11-*cis*-retinal, acts as an inverse agonist in the rhodopsin ground state. The capture of a photon by rhodopsin induces the isomerization of 11-*cis*-retinal into all-*trans*-retinal approximately in 200 femtoseconds (Figure1.3). The structural perturbation in the retinal-binding pocket and TM helical bundle drives the passage through a series of spectrally different photointermediates in picosecond to millisecond time scales. These intermediates are characterized by their distinct absorption maxima. The Schiff base is protonated in dark state of rhodopsin and as well as the intermediates until Meta I. The protonated shiff base can be understood from their absorption maxima being greater than 380nm. Relaxations in the protein and alterations in the protein-chromophore interactions ultimately lead to tertiary structure changes that are the hall-mark of the active state of rhodopsin, metarhodopsin II (Meta II). Meta II is distinguished by an unprotonated Schiff base imine and a dramatically blue-shifted λ_{max} value (380 nm) in the absorption curve (Figure1.4). Within minutes, Meta-II decays to free all-*trans*-retinal and opsin apoprotein by the irreversible hydrolysis of the Schiff base linkage. After all-*trans*-retinal is reduced to retinol and 11-*cis*-retinal is regenerated in the neighboring retinal pigment epithelial cells, opsin spontaneously combines with 11-*cis*-retinal to form rhodopsin (Wald, 1952;Wald, 1953).

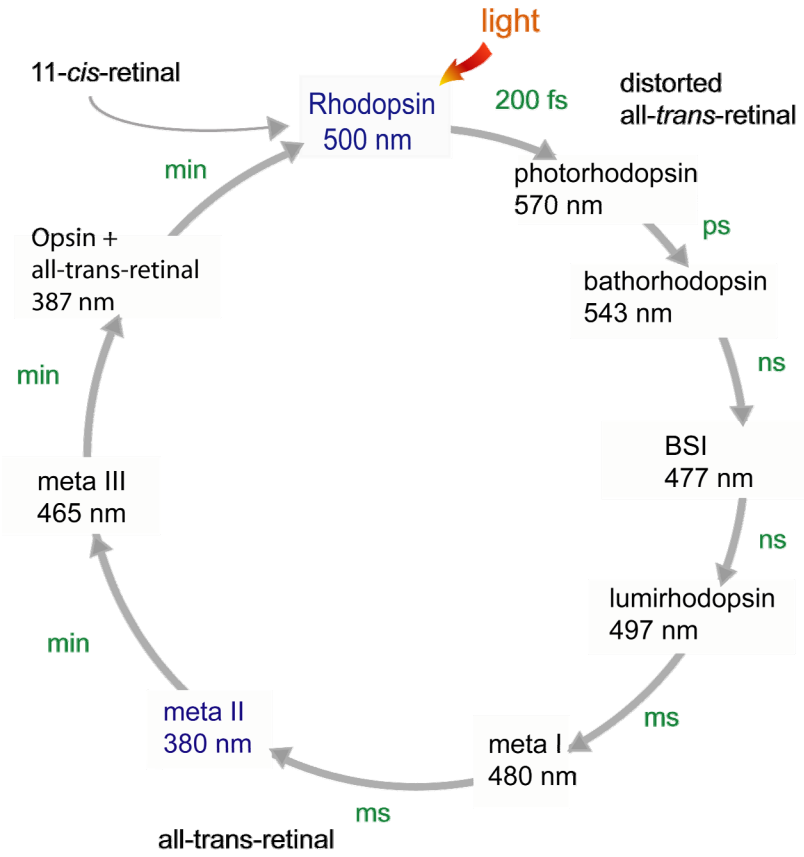

Figure 1.4 Photocycle of rhodopsin with spectroscopically detected intermediates. Absorption of light by rhodopsin leads to photoisomerization of the 11-*cis* double bond on a femtosecond time scale. This isomerization is one of the fastest photochemical events that happen in nature. The first photoproduct, photorhodopsin, has a highly distorted 11-*trans* bond. Then, photolyzed pigment thermally relaxes through multiple spectral inactive intermediates which are determined by their distinct λ_{max} values by low-temperature spectroscopic studies or flash photolysis at room temperature. Distinct λ_{max} values of these intermediates stem from gradual conformational changes in the protein as well as protein-chromophore interactions. Up to Meta I, protonation of the Schiff base persists due to stabilizing effect of the counterion Glu113. Deprotonation of the Schiff base results in a large blue-shifted value for λ_{max} (380 nm) specific to Meta II state which is in dynamic equilibrium with Meta I under physiological conditions. Meta II is the active state with larger confomatinal changes in the protein and all-*trans*-retinal. It is the only photoproduct which is capable of binding and activating the G-protein–transducin leading to catalysis guanine nucleotide exchange. Meta II eventually decays to free all-*trans*-retinal and opsin apoprotein with irreversible hydrolysis of the Schiff base linkage. Meta II decay can occur either directly or through a species termed Meta III with a reprotonated retinal Schiff base (λ_{max}=465 nm). Then, the free all-*trans*-retinal is reduced to retinol by retinol dehydrogenase and transported to the retinal pigment epithelial cells where 11-*cis*-retinal is regenerated. Later, 11-*cis*-retinal carried back to photoreceptor cells and opsin combines (Wald, 1952;Wald, 1953).

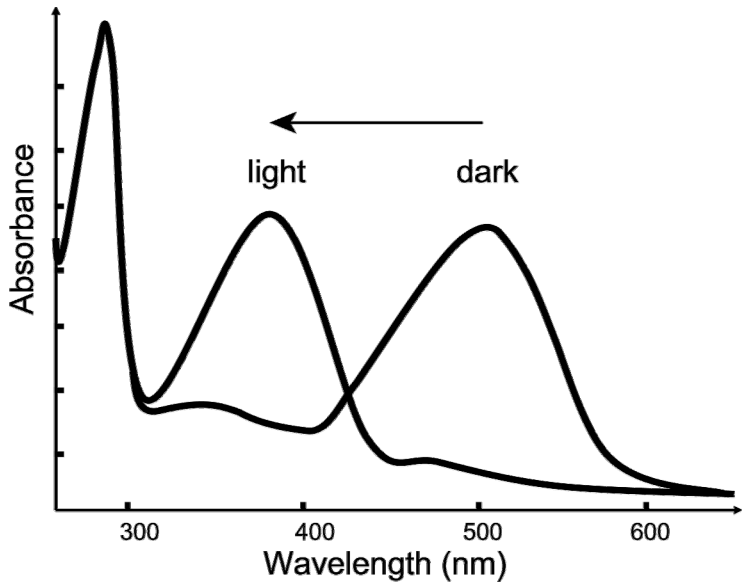

Figure 1.5 Ultraviolet-Visible absorption spectrum of rhodopsin and Meta II state.
The characteristic absorption spectrum of rhodopsin (dark) shows a broad visible absorbance with a maximum wavelength (λ_{max}) of 500 nm. This peak originates from the interaction between the chromophore at 11-*cis*-retinal form and its binding pocket in the apoprotein opsin. The 280 nm peak represents the protein component. Within milliseconds after exposure to light at room temperature, the G-protein-interacting active state, metarhodopsin II, is reached with a λ_{max} value of 380 nm (image taken from Sakmar et al., 2002)

1.2.3 Rhodopsin activation

Current knowledge on the activation mechanism of rhodopsin has been acquired from a range of biophysical and biochemical experiments with various degrees of detail. These include site-directed mutagenesis, photoactivity, NMR, and engineered disulfide bonds to rhodopsin and other GPCRs (Meng and Bourne, 2001;Klein-Seetharaman, 2002;Hubbell et al., 2003). Such studies of the dark and light-adapted states of rhodopsin have suggested that the activation process is coupled to rigid-body movements of the TM helices (Hubbell et al., 2003). Accordingly, a conserved D(E)RY motif and other amino acids that interact with transducin are exposed via translations of H2, H3, H6, and H7, and rotation of H6. The term “rigid-body”

indicates that the helices maintain their secondary structures, although their relative positions and orientations in space undergo displacements (Klein-Seetharaman, 2002). The fluctuations are highest in the loops connecting the helices, as evidenced by the high flexibility of spin labels introduced in the CP loops, the formation of disulfide bonds between engineered cysteines located approximately 5-6 Å apart within the CP surface and the high B-factors observed in loop regions (Klein-Seetharaman, 2002;Hubbell et al., 2003). The movements of the helices during light-activation presumably involve the disruption of specific interhelical contacts and the release of the associated constraints. Despite the large body of literature on rhodopsin activation, the details of what drives these motions and the molecular interactions that underlie the experimental data on the Meta II state remain unknown. Epitomizing this lack of molecular understanding, the fluorescence decay rates observed for various rhodopsin mutants during Meta II deactivation are generally treated as "black-box" values (Farrens and Khorana, 1995).

Recently, a photoactivated intermediate of rhodopsin has been crystallized at a resolution of 4.15 Å. In this state, the CP loops connecting helices were disordered due to high flexibility as also evidenced by spin labels introduced in these regions (Farahbakhsh et al., 1995;Farrens et al., 1996;Altenbach et al., 1996). However, in critical regions at the CP site, conformational changes predicted by the photoactivated crystal structure are much smaller than those found by biophysical and biochemical experiments (Menon et al., 2001;Sakmar et al., 2002;Hubbell et al., 2003). Locations of all backbone atoms are not exactly determined, most side-chain and chromophore coordinates absent in Protein Data Bank (PDB) (Berman et al., 2000) structure (PDB code 2I37) (Salom et al., 2006) of this photoactivated state. Further studies have been suggested to determine the relationship of this photoactivated intermediate in the crystals to Meta II (Salom et al., 2006).

These studies have shown that conformational changes on and near the CP domain of rhodopsin are critical to formation of active rhodopsin and the G-protein binding domain on this site. Movements of helices presumably involve the breakage of specific contacts, and the release of the associated constraints. These constraints cluster in several conserved "microdomains": (i) an ionic interaction between charges of ligand and receptor corresponding to the retinal Schiff base and the Glu113 counterion in rhodopsin (Sakmar et al., 1991;Cohen et al., 1993), (ii) highly conserved D(E)RY motif at the CP end of H3 and the $X_1BBX_2X_3B$ motif at the CP end of H6 (B, basic; X, non-basic) (Samama et al., 1993;Hogger et al., 1995;Scheer et al., 1996;Ballesteros et al., 1998;Rasmussen et al., 1999;Alewijnse et al., 2000;Ballesteros et al., 2001;Visiers et al., 2002;Greasley et al., 2002), (iii) the Asn-Asp pair at H1 and H2, respectively, and the NPXXY motif at H7 (Okada et al., 2002) and (iv) the aromatic cluster surrounding the ligand binding pockets (Nakayama and Khorana, 1990;Nakayama and Khorana, 1991;Borhan et al., 2000).

1.2.4 Computational efforts for understanding the dynamics of rhodopsin.

For different states of rhodopsin, experimental studies provide valuable and extensive information on the surface accessibility of certain residues, the relative orientation of helices and, the interactions of the retinal with residues of the chromophore binding pocket; however the mechanism of activation and the active structure of rhodopsin still remain unknown. Recent advances in efficient modeling and simulation of protein dynamics lead many computational analyses of the rhodopsin dark structure that help to elucidate the activation mechanism of rhodopsin. The first MD simulations of photoisomerization process were performed by Rohrig et al. (2002) and by Schulten and coworkers (2002). These simulations provided information on the

early events of the opsin relaxation in response to light activation. More recent simulations spanning time scales up to 15 ns focused on the effects of lipids (Crozier et al., 2003;Huber et al., 2004), the time evolution of the chromophore isomerization accompanying the transition to the luminorhodopsin, (Lemaitre et al., 2005) and the coupling between retinal and larger conformational changes (Crozier et al., 2003;Crozier et al., 2007). However, the rhodopsin to Meta II transition, occurring in milliseconds, remains beyond the reach of MD simulations that are usually confined to timescales on the order of tens of nanoseconds. Starting from a bathorhodopsin model, (Ishiguro et al., 2003) structural models have been proposed for later photointermediates, including lumirhodopsin, Meta I and Meta II by swinging the C- and N-terminal ends of H3 and rotating H6 (Ishiguro et al., 2004). In other studies, distance constraints derived from NMR and EPR spectroscopy and/or disulfide bond formation experiments have been used to model the Meta II structure (Choi et al., 2002;Gouldson et al., 2004;Niv et al., 2006). In another approach results from energy calculations for the TM region and experimental data have been iteratively combined to generate a model for Meta II (Nikiforovich and Marshall, 2003), assuming the non-TM fragments to retain similar conformations in the dark and the light activated states. However, it is known from a large number of experiments that these regions, in particular the CP loops and the soluble H8, are highly mobile and their mobilities are important for activation (Klein-Seetharaman, 2002). Later, Marshall and coworkers built sets of low energy conformers for loops of rhodopsin both in the dark and the light activated states by using TM regions of their previous models as templates (Nikiforovich and Marshall, 2005). Effects of unsaturated lipids on rhodopsin stability and kinetics were studied by analyzing 26 independent 100 ns simulations of dark-adapted rhodopsin and results showed that the presence of tightly packed unsaturated lipids tends to weaken the interhelical packing by making specific interactions with the protein (Grossfield et al., 2006). The same group also used these trajectories

to study the sampling efficiency of MD simulations of rhodopsin. Analyses showed that fluctuations of rhodopsin are not well described in these simulations especially for the helical bundle (Grossfield et al., 2007).

1.2.5 Interaction of rhodopsin with other proteins of the visual signaling cascade

1.2.5.1 Meta II activates multiple transducin molecules

The Meta II form has the ability to associate with the heterotrimeric G-protein, transducin. Inactive form of transducin is composed of αβγ subunits ($G_{\alpha\beta\gamma}$). Binding of Meta II promotes activation and exchange of guanosine diphosphate (GDP) to guanosine triphosphate (GTP) in $G_{\alpha\beta\gamma}$. Then, tightly bound G_{α} and $G_{\beta\gamma}$ subunits separate from each other and from the receptor as well. Both components are free to interact and modulate the activity of downstream elements of the visual signaling cascade. Each Meta II can activate hundreds of transducins and this is the first amplification step in the cascade.

1.2.5.2 Rhodopsin kinase and arrestin leads to signal quenching

Since the Meta II decay rates takes minutes, CP proteins such as rhodopsin kinase and arrestin interact with Meta II for signal quenching and light adaptation. Rhodopsin kinase phosphorylates Meta II at Ser334, Ser338 and Ser343. Arrestin binds phosphorylated Meta II to prevent it from activating transducin (Hubbell et al., 2003).

1.3 OPEN QUESTIONS

Even with a wide range of information available on both dark and active state of rhodopsin, details about how the conformational changes necessary for activation are triggered and the molecular mechanisms explaining the experimental data on the Meta II state still remain unknown. Furthermore, the current time scale explored in conventional simulation techniques such as MD are too short for describing long time scale motions of proteins. The following questions that are relevant to the activation mechanism of rhodopsin and to the computational techniques for studying protein dynamics were raised here:

What are the conformations of rhodopsin which couple retinal isomerization to the conformational changes in both the TM domain and the critical G-protein binding sites on the CP surface? Are these allosteric motions intrinsically built-into the structure such that they can be elucidated by coarse grained elastic models? If yes, what would be the activation model for rhodopsin derived from the global modes of GNM and ANM? How the model can be further refined by including residue specificities and, the lipid and water molecules around the protein? How can we study the relevance of multiple ANM modes to the biological function of the protein? What would be the direction and scale of the deformation for each mode? How could MD and coarse grained models be combined to observe large scale motions of proteins?

1.4 SPECIFIC AIMS AND SUMMARY OF ACCOMPLISHMENTS

1.4.1 Specific aims

1.4.1.1 Specific aim 1. Understanding the activation mechanism of rhodopsin by methodologies with different levels of "specificity" and refining the results with experimental findings.

First, we aim at finding the intrinsic motions of rhodopsin encoded by its structure, upon modeling it as a coarse-grained ENM. Then, we explore the relevance of these motions to biological function. The resulting conformations obtained from the analysis of GNM/ANM modes are compared to experimental data on activation including site-directed mutagenesis, cysteine cross-linking, and site-directed spin labeling analyses. We propose an activation model that elucidates the coupling of retinal isomerization to conformational changes in both the TM domain and the critical G-protein binding sites on the CP surface. Additionally, we identify the structural elements that stabilize rhodopsin using FIRST software which simulates the thermal unfolding of a protein by breaking the native-state hydrogen bonds sequentially in the order of their relative strength.

While lacking details, the motions captured by the ANM provide a good estimate of collective dynamics beyond the range of conventional MD simulations. To further refine the activation model proposed by GNM/ANM analysis and to study the effects of environment on rhodopsin dynamics, we use the newly developed method in Specific Aim 2.

1.4.1.2 Specific aim 2. Developing a new method of expanded capacity to efficiently model biomolecular motions and applying it to rhodopsin.

To this aim, we propose an algorithm referred as **ANM-restrained MD**, which uses the deformations derived from ANM analysis as restraints in MD trajectories. We restrict the degrees of freedom of the system to those that are critical to determine a conformational change relevant to its biological function. This permits us to sample the collective motions that are otherwise beyond the range of conventional MD simulations. By this new approach, we aim at incorporating the realism and accuracy of MD into ANM analysis while taking advantage of ANM results to accelerate MD simulations. The ANM-restrained MD protocol enables us to observe the cooperative changes at full atomic detail in the presence of the environment, i.e. explicit water and membrane molecules (Isin et al., 2008).

Next, we will apply **ANM-restrained MD** algorithm to rhodopsin. We would like to explore the contribution of multiple modes to the dynamics of rhodopsin, while incorporating the effects from residue specificities and from interactions with lipid and water molecules.

1.4.2 Summary of accomplishments

1.4.2.1 ANM-restrained MD method

We developed the following protocol to use collective modes of ANM as guides in MD simulation. The approach allows for a search in the conformational space by constraining the system only minimally to global motions of the protein (Isin et al., 2008).

1. Normal modes are generated using ANM. A subset of modes that are sufficiently decoupled from others, based on the eigenvalue distribution, is selected.

2. Starting from the first mode associated with the lowest eigenvalue, harmonic restraints are applied in two opposite directions (plus and minus) in MD simulations.
3. The resulting two conformations are subjected to energy minimization to relieve possible unrealistic distortions lead by the restraints. The conformer with the lower energy is then selected as the starting structure for the application of the next mode as new harmonic restraints.
4. When all modes in the subset are utilized in MD, a new set of modes is generated by ANM and the procedure described above is repeated using the new subset of modes.

1.4.2.2 Insights about global motions of rhodopsin

First, we found the cooperative global motions of rhodopsin by using coarse grained ENMs, GNM and ANM. To explore the biological relevance of these motions to the activation of rhodopsin, we did an extensive comparison with the experimental data available on rhodopsin activation and stability. Next, we used the FIRST software in conjunction with the high frequency modes of GNM to find the residues that are important for functional fold.

Then, following ANM-restrained MD protocol, we added the effects of multiple global modes, residue specificity, and the environment on rhodopsin dynamics simulations. We identified a hinge site that does not change upon several rounds of applying normal modes as restraints. It includes residues that are directly affected by the isomerization of retinal and consists of Ala124 and Leu125 at H3; Trp161 and Cys167 at H4; Pro180, Glu181, Ser186, Cys187 and Gly188 on the β-strands at EL2; Met207, His211, Phe212 at H5; Trp265, Tyr268 at H6; and Tyr292 and Ala295 at H7. This hinge contains residues that define the 11-*cis*-retinal binding pocket at the dark state as well as those stabilizing all-*trans*-retinal at the structure

reached by the end of our simulations. The chromophore binding pocket is a tightly packed environment. Hence, the hinge site residues are sensitive to small conformational changes in retinal and directly participate in rhodopsin activation.

In the dark state structure, H6 is highly constrained by 11-*cis*-retinal. Aromatic ring of Trp265 makes several contacts with β-ionone ring of 11-*cis*-retinal. Our results show that the isomerization and the flip of β-ionone ring release constraints imposed at H6 and lead to conformation changes. In contrast, H4 and H5 form new contacts with all-trans-retinal and the number of contacts between H3 residues and all-*trans*-retinal changes. Initially, 11-*cis*-retinal was in close contact with H6 residues such as Phe261, Trp265, Tyr268 and Ala269. At the end of our simulations, the all-*trans*-retinal made new atomic contacts including Cys167 at H4, Phe203 at H5 which were not accessible in the dark state. The number of atom-atom contacts with Met207 and His211 increased in agreement with NMR experiments (Patel et al., 2004). In particular, the new contact found between Cys167 at H4 and the all-*trans*-retinal support our experiments showing the significant effect of Cys167Ser on Meta II decay rate (Isin et al., 2006).

Rearrangement of helices to accommodate the changes in the retinal binding site leads to opening of the helical bundle at the CP end and the exposure of the helical bundle. The CP ends of the helices H3, H4, H5, and H6 and connecting loops are found enjoy an enhanced mobility facilitated by this hinge site.

We included water molecules in our simulations by using the newly proposed protocol, and found another stable region close to the CP ends of H1, H2 and H7. At this region, two water molecules were found be connected to highly conserved residues throughout the simulations.

Two residues belonging to the NPXXY motif at H7, Asn302 and Tyr306, are connected to H1 and H2 through two water molecules located at the cavity between helices H1, H2 and H7. At the structure reached by the end of our simulations, Asn55 (H1), Asp83 (H2), and Asn302 (H7) are hydrogen bonded to a central water molecule; and a second water molecule interacts closely with Thr62 (H1), Asn73 (H2), and Tyr306 (H7). This second water molecule was already found to be interacting with these residues by X-ray crystallography and still exists at the same vicinity.

2.0 PREDISPOSITION OF THE DARK STATE OF RHODOPSIN TO FUNCTIONAL CHANGES IN STRUCTURE

2.1 ABSTRACT

A model for the Meta II state is proposed here by analyzing elastic network normal modes in conjunction with experimental data. Key mechanical features and interactions broken/formed in the proposed model are found to be consistent with the experimental data. We further tested the model by using a set of Meta II fluorescence decay rates measured to empirically characterize the deactivation of rhodopsin mutants. We find that our model correctly predicts 93% of the experimentally observed qualitative response in 119 rhodopsin mutants for which the decay rates and misfolding data have been measured, including a systematic analysis of Cys->Ser replacements reported here. Based on the detailed comparison between model and experiments, a cooperative activation mechanism is deduced which couples retinal isomerization to concerted changes in conformation, facilitated by the intrinsic dynamics of rhodopsin. We identify a global hinge site near the retinal-binding pocket that ensures the efficient propagation of signals from the central TM region to both CP and EC ends. The predicted activation mechanism opens the TM helices at the critical G-protein binding CP domain. This model provides a detailed, mechanistic description of the activation process, extending experimental observations and yielding new insights for further tests.

2.2 INTRODUCTION

In the absence of a three-dimensional (3D) structure for the Meta II state, current knowledge on the activation mechanism of rhodopsin has been acquired from a range of biophysical and biochemical experiments with various degrees of detail. Due to recent advances in efficient modeling and simulation of protein dynamics, there has been a flourish of computational analyses of the rhodopsin dark structure to suggest a structural model for the active state. The large scale motions required to fit experimental data suggest that other, faster computational methods are needed to explore the collective activation mechanism of rhodopsin (Rader et al., 2004). Our aim is to gain a better understanding of molecular mechanisms that drive the transition to the Meta II state by integrating experimental and computational results.

In this study, we present an in-depth analysis of the slow modes of motions predicted by the GNM/ANM, to characterize the global dynamics of rhodopsin in the dark state and identify mechanisms amenable to its activation. We combine these results to create an active, Meta II state model, which is refined and validated by comparisons to a variety of experimental data. We show that the opening of the helical bundle in the CP domain is favored by the intrinsic dynamic properties of rhodopsin, as a consequence of the particular topology of inter-residue contacts in the dark state. Specifically, we identify a hinge site near the retinal binding pocket and show how this region coordinates the cooperative propagation of structural changes to the CP region. The opening of the helical bundle at the G-protein binding CP site is shown to be controlled by the concerted motion of the amino acids near the ligand binding pocket, providing a molecular explanation for structural changes observed upon formation of the active Meta II state, and a molecular interpretation of Meta II decay rates and disulfide bridge formation data.

2.3 METHODS

2.3.1 Elastic Network Models

In the elastic network model of a protein (Tirion, 1996; Bahar et al, 1997), the interactions between residues in close proximity are simplified by purely harmonic potentials. In particular, the nodes are residues identified by the C^{α} atoms. No distinction is made between different types of amino acids and a generic force constant is adopted for the inter-residue potential. Any deformation from the native state coordinates is resisted by linear springs, associated with closely neighboring residues and representing bonded and non-bonded interactions (Figure 2.1). The physical basis and the analytical method for examining the collective dynamics of such a mass-spring network, bears close similarities to the elasticity theory of random polymer networks (Flory, 1976). To this end, we will discuss below two different elastic network models, Gaussian and Anisotropic networks for modeling protein equilibrium dynamics.

2.3.1.1 Gaussian Network Model

In the GNM (Bahar et al., 1997; Haliloglu et al., 1997), the residues are assumed to undergo Gaussianly distributed fluctuations around their mean positions, coupled by harmonic potentials. To study the dynamics of such a network, we define the equilibrium position of a node *i* by the vector $\mathbf{R}_i^0$ and its instantaneous position by $\mathbf{R}_i$ (Figure 2.1A). The fluctuation from the mean position is given by $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$. For a pair of interacting residues *i* and *j,* we denote the equilibrium inter-node distance vector extending from residue *i* to residue *j* as $\mathbf{R}_{ij}^0$ and the

instantaneous separation vector as $\mathbf{R}_{ij}$. Accordingly, the fluctuation in the inter-node distance vector can be written as $\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^{0} = \Delta\mathbf{R}_{j} - \Delta\mathbf{R}_{i}$.

In the GNM, these fluctuations are assumed to be *isotropic* obeying the distribution

$$\mathrm{W}\left(\Delta\mathbf{R}_{ij}\right) \propto \exp\left\{\frac{-\gamma}{2k_{B}T}\Delta\mathbf{R}_{ij} \bullet \Delta\mathbf{R}_{ij}\right\} \tag{2.1}$$

where k_B is the Boltzmann constant, T is the absolute temperature and $\Delta\mathbf{R}_{ij} \bullet \Delta\mathbf{R}_{ij}$ is the dot product denoting the square of the fluctuation magnitude.

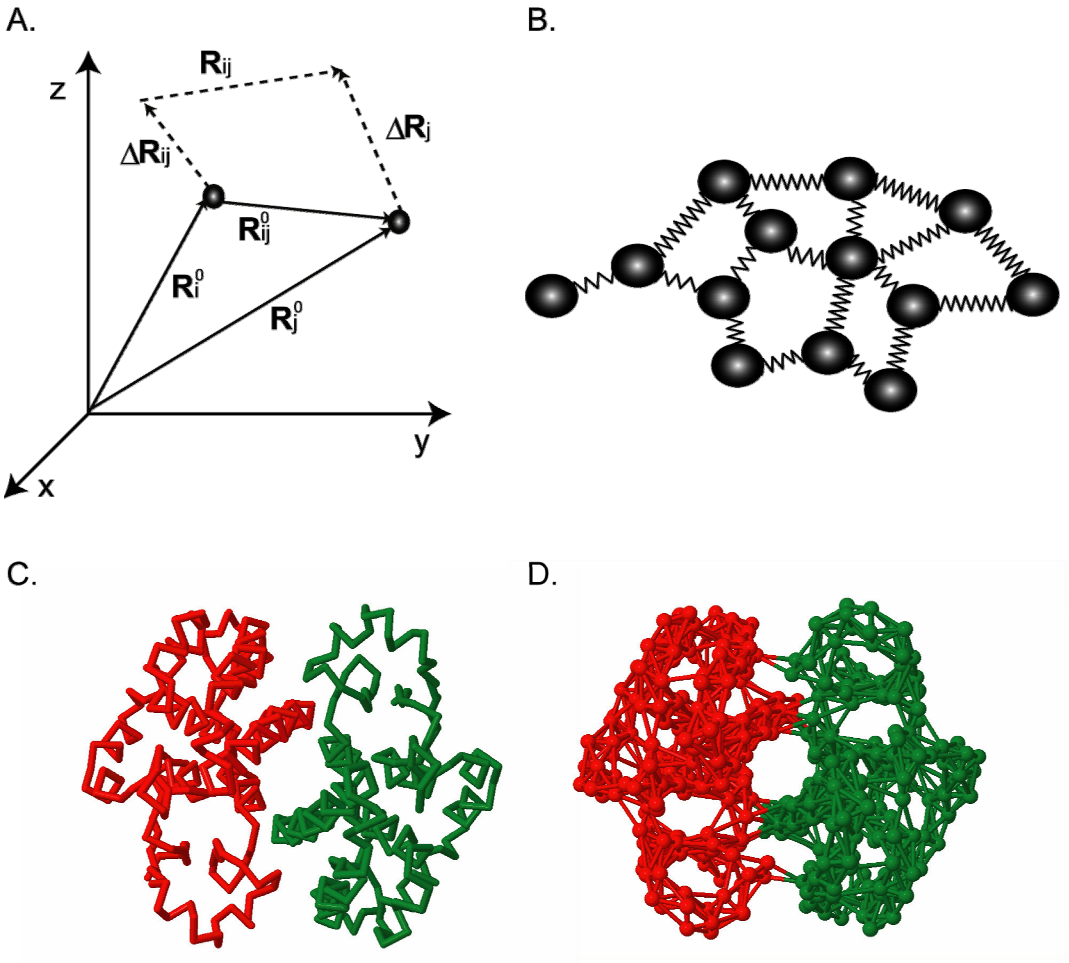

Figure 2.1 Representations in GNM and ANM

(A) Schematic representation of the equilibrium positions of the i^{th} and j^{th} nodes, $\mathbf{R}_i^0$ and $\mathbf{R}_j^0$, with respect to the coordinate system (xyz). Their instantaneous fluctuation vectors $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$, are shown by the dashed arrows, as well as the instantaneous separation vector $\mathbf{R}_{ij}$ between the positions of the two residues. $\mathbf{R}_{ij}^0$ is the equilibrium distance between nodes *i* and *j*. (B) In the GNM, every residue is represented by a node and connected to spatial neighbors by uniform springs. These springs determine the N-1 degrees of freedom in the network and the structure's modes of vibration. (C) Three dimensional image of glutathione transferase (PDB code 2A2R) (Tellez-Sanz et al., 2006) showing the C^α-trace. (D) Using a cutoff value of 10Å, all connections between C^α nodes are drawn for glutathione transferase structure to indicate the nature of the elastic network analyzed by GNM and ANM.

For the probability density given in Eq. (2.1), we can infer the equivalent interaction potential between residues *i* and *j* to be $\frac{\gamma}{2}\mathbf{\Delta R}_{ij} \bullet \mathbf{\Delta R}_{ij}$. Physically, this means that both the changes in the inter-residue distances, and changes in the direction of the inter-residue distance vector, are being penalized. The potential can be expressed in terms of the x-, y- and z-components of the fluctuation vector is:

$$V_{GNM}^{ij} = \frac{\gamma}{2}\left[\left(\Delta \mathrm{X}_i - \Delta \mathrm{X}_j\right)^2 + \left(\Delta \mathrm{Y}_i - \Delta \mathrm{Y}_j\right)^2 + \left(\Delta \mathrm{Z}_i - \Delta \mathrm{Z}_j\right)^2\right] \tag{2.2}$$

The potential can be extended over *all* pairs of interactions in a network of *N* residues

$$V_{GNM} = \frac{\gamma}{2}\left[\sum_{i,j}^{N} \Gamma_{ij}\left\{\left(\Delta \mathrm{X}_i - \Delta \mathrm{X}_j\right)^2 + \left(\Delta \mathrm{Y}_i - \Delta \mathrm{Y}_j\right)^2 + \left(\Delta \mathrm{Z}_i - \Delta \mathrm{Z}_j\right)^2\right\}\right] \tag{2.3}$$

where Γ_{ij} is the ijth element of a Kirchhoff matrix of inter-residue contacts defined by

$$\Gamma_{ij} = \begin{cases} -1 & if \quad i \neq j \quad and \quad R_{ij}^0 \leq r_c \\ 0 & if \quad i \neq j \quad and \quad R_{ij}^0 > r_c \\ -\sum_{i \neq j} \Gamma_{ij} & if \quad i = j \end{cases} \tag{2.4}$$

and R_{ij}^0 is the magnitude of the equilibrium distance vector $\mathbf{R}_{ij}^0$. $\mathrm{r_c}$ is the cutoff distance that defines the connected residues. A lower bound for r_c is the first interaction shell radius of 7.0 Å extracted from statistical analyses (Miyazawa and Jernigan, 1985, 1996; Bahar and Jernigan, 1997) of PDB structures. Simplifying further, the overall potential defined from GNM is

$$V_{GNM} = \frac{\gamma}{2}\left[\mathbf{\Delta X}^T \mathbf{\Gamma \Delta X} + \mathbf{\Delta Y}^T \mathbf{\Gamma \Delta Y} + \mathbf{\Delta Z}^T \mathbf{\Gamma \Delta Z}\right] \tag{2.5}$$

where $\mathbf{\Delta X}^T$, $\mathbf{\Delta Y}^T$ and $\mathbf{\Delta Z}^T$ are the row vectors $[\Delta \mathrm{X}_1 \;\; \Delta \mathrm{X}_2 \;\; ..... \;\; \Delta \mathrm{X}_N]$, $[\Delta \mathrm{Y}_1 \;\; \Delta \mathrm{Y}_2 \;\; ..... \;\; \Delta \mathrm{Y}_N]$ and $[\Delta \mathrm{Z}_1 \;\; \Delta \mathrm{Z}_2 \;\; ..... \;\; \Delta \mathrm{Z}_N]$, respectively, with the superscript *T* denoting the transpose. Following the

isotropic assumption of GNM, the fluctuations in x-, y- and z- directions are $\left\langle (\Delta \mathbf{X})^2 \right\rangle = \left\langle (\Delta \mathbf{Y})^2 \right\rangle = \left\langle (\Delta \mathbf{Z})^2 \right\rangle = \left\langle (\Delta \mathbf{R})^2 \right\rangle / 3$.

The potential is given in equation 2.5 gives rise to a probability distribution over fluctuations $\mathbf{\Delta R}$ that is Gaussian,

$$\mathrm{W}(\mathbf{\Delta R}) = \mathrm{W}(\mathbf{\Delta X})\,\mathrm{W}(\mathbf{\Delta Y})\,\mathrm{W}(\mathbf{\Delta Z}) \tag{2.6}$$

with

$$\mathrm{W}(\mathbf{\Delta X}) = \frac{1}{Z_{\mathbf{\Delta X}}} \exp\left\{ \frac{-\gamma}{2k_B T} \mathbf{\Delta X \Gamma \Delta X} \right\} \tag{2.7}$$

The partition function $Z_{\mathbf{\Delta X}}$ associated with the fluctuations along the *x*-direction is given by the determinant

$$Z_{\mathbf{\Delta X}} = (2\pi)^{\frac{(N-1)}{2}} \left(det\left[\frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right] \right)^{1/2} \tag{2.8}$$

By definition, $\mathbf{\Gamma}$ is positive semi-definite, i.e. $\mathbf{\Gamma}$ has rank *N-1*, so it cannot be inverted. Instead, its pseudo-inverse is computed after eliminating the contribution from the zero eigenvalue. Because of the isotropic assumption, similar expressions can be written for $\mathrm{W}(\mathbf{\Delta Y})$ and $\mathrm{W}(\mathbf{\Delta Z})$. So a configurational partition function for the GNM can be expressed, in analogy with the theory of random networks in polymer science (Flory 1976), as

$$Z_{\mathbf{\Delta R}} = Z_{\mathbf{\Delta X}} Z_{\mathbf{\Delta Y}} Z_{\mathbf{\Delta Z}} = (2\pi)^{\frac{3(N-1)}{2}} \left(det\left[\frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right] \right)^{3/2} \tag{2.9}$$

From the Gaussian distribution, we can infer the equilibrium cross-correlations in residue fluctuations from

$$\left\langle \Delta\mathbf{R}_i \bullet \Delta\mathbf{R}_j \right\rangle = \frac{1}{Z_{\Delta\mathbf{R}}} \int \left(\Delta\mathbf{R}_i \bullet \Delta\mathbf{R}_j \right) e^{-V_{GNM}/k_B T} d\Delta\mathbf{R} = \left(3k_B T/\gamma \right) \left[\mathbf{\Gamma}^{-1} \right]_{ij} \tag{2.10}$$

where $\left[\mathbf{\Gamma}^{-1} \right]_{ij}$ is the ij^{th} element of the inverse of $\mathbf{\Gamma}$. The mean square (ms) fluctuations $\left\langle \left(\Delta\mathbf{R}_i \right)^2 \right\rangle$ of individual residues are evaluated from the last equality in equation 2.10, using $i = j$.

The eigenvalue decomposition of $\mathbf{\Gamma}$ reads $\mathbf{\Gamma} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^{\mathbf{T}}$ where $\mathbf{U}$ is an orthogonal matrix whose columns $\mathbf{u}_i$, $1 \le i \le N$, are the eigenvectors, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues λ), usually organized in an ascending order, i.e., $\lambda_1 = 0 \le \lambda_i \le \lambda_N$. The i^{th} eigenvector reflects the *shape* of the i^{th} mode as a function of residue index, and the i^{th} eigenvalue represents its *frequency* (Haliloglu et al., 1997; Bahar and Jernigan, 1998; Bahar, 1999). The correlations between the fluctuations of residues i and j result from the superposition of $(N\text{-}1)$ modes in the GNM. The contribution of the k^{th} mode, $\left\langle \mathbf{\Delta R}_i \cdot \mathbf{\Delta R}_j \right\rangle_k$, is given by

$$\left\langle \mathbf{\Delta R}_i \cdot \mathbf{\Delta R}_j \right\rangle_k = (3\mathrm{k_B T}/\gamma) \left[\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^{\mathrm{T}} \right]_{ij} \tag{2.11}$$

and the cumulative effect of a range of modes $k_1 \le k \le k_2$ is found by summing up the above equation over the investigated subset.

2.3.1.2 Anisotropic Network Model: Mechanisms of collective motions

The *mechanisms* of motions are determined by an extension of the GNM to the *3N-d* space of collective modes, using the ANM (Doruker et al., 2000; Atilgan et al. 2001). Whereas the inter-residue 'distances' are controlled by harmonic potentials in the GNM, ANM adopts the

further assumption that the three (x-, y- and z-) components of the inter-residue separation vectors obey each Gaussian dynamics. ANM thus involves the inversion of a *3N* x *3N* Hessian matrix **H** that replaces the *N* x *N* Kirchhoff matrix **Γ** by

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N} \\ \mathbf{H}_{21} & \ddots & \vdots \\ \vdots & & \\ \mathbf{H}_{N1} & \cdots & \mathbf{H}_{NN} \end{pmatrix} \tag{2.12}$$

H is composed of *N* x *N* super-elements $\mathbf{H}_{ij}$ ($1 \le i, j \le N$) each of size 3 x 3 and given by

$$\mathbf{H}_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial \mathrm{X}_i \partial \mathrm{X}_j} & \frac{\partial^2 V}{\partial \mathrm{X}_i \partial \mathrm{Y}_j} & \frac{\partial^2 V}{\partial \mathrm{X}_i \partial \mathrm{Z}_j} \\ \frac{\partial^2 V}{\partial \mathrm{Y}_i \partial \mathrm{X}_j} & \frac{\partial^2 V}{\partial \mathrm{Y}_i \partial \mathrm{Y}_j} & \frac{\partial^2 V}{\partial \mathrm{Y}_i \partial \mathrm{Z}_j} \\ \frac{\partial^2 V}{\partial \mathrm{Z}_i \partial \mathrm{X}_j} & \frac{\partial^2 V}{\partial \mathrm{Z}_i \partial \mathrm{Y}_j} & \frac{\partial^2 V}{\partial \mathrm{Z}_i \partial \mathrm{Z}_j} \end{bmatrix} \tag{2.13}$$

ANM is a simplified version of NMA and superelements of the **H** matrix can be derived from the second derivatives of the potential function *V*. If we consider residues *i* and *j*, the harmonic potential for them will be defined as

$$\begin{aligned} V_{ANM}^{ij} &= \frac{\gamma}{2}\left[R_{ij} - R_{ij}^0\right]^2 \\ &= \frac{\gamma}{2}\left(\left[\left(\mathrm{X}_j - \mathrm{X}_i\right)^2 + \left(\mathrm{Y}_j - \mathrm{Y}_i\right)^2 + \left(\mathrm{Z}_j - \mathrm{Z}_i\right)^2\right]^{1/2} - R_{ij}^0\right)^2 \end{aligned} \tag{2.14}$$

where R_{ij}^0 and R_{ij} are the equilibrium and instantaneous separation *distances* between sites *i* and *j*, respectively. The overall potential is given by

$$V_{ANM} = \sum_{i,j} \frac{\gamma}{2} \left[R_{ij} - R_{ij}^0 \right]^2 \Theta\left(r_c - R_{ij} \right) \tag{2.15}$$

where $\Theta\left(r_c - R_{ij} \right)$ is the Heavyside step function equal to 1 if the argument is positive and zero otherwise. $\Theta\left(r_c - R_{ij} \right)$ selects those residue pairs within a cut-off distance of r_c. The cut-off radius is typically larger in ANM (about 15 Å) than GNM (about 10Å) (Atilgan et al., 2001).

The first and second derivatives of V_{ANM} with respect to the components of the position vectors $\mathbf{R}_i$ and $\mathbf{R}_j$ read

$$\frac{\partial V_{ANM}}{\partial \mathrm{X}_i} = - \frac{\partial V_{ANM}}{\partial \mathrm{X}_j} \tag{2.16}$$

$$= -\gamma \left(\mathrm{X}_j - \mathrm{X}_i \right) \left(1 - \frac{R_{ij}^0}{R_{ij}} \right) \tag{2.17}$$

and

$$\frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_i^2} = - \frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_j^2} \tag{2.18}$$

$$= -\gamma \left(1 + \left(\mathrm{X}_j - \mathrm{X}_i \right)^2 \frac{R_{ij}^0}{R_{ij}^3} - \frac{R_{ij}^0}{R_{ij}} \right) \tag{2.19}$$

respectively.

At equilibrium, $R_{ij}^0 = R_{ij}$ and the equations 2.17 and 2.19 simplify to

$$\frac{\partial V_{ANM}}{\partial \mathrm{X}_i} = 0 \tag{2.20}$$

$$\frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_i^2} = \frac{\gamma}{R_{ij}^2}\left(\mathrm{X}_j - \mathrm{X}_i\right)^2 \tag{2.21}$$

The second cross-derivative can be written as

$$\frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_i \partial \mathrm{Y}_j} = -\frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_j \partial \mathrm{Y}_i} \tag{2.22}$$

$$= -\gamma \frac{\left(\mathrm{X}_j - \mathrm{X}_i\right)\left(\mathrm{Y}_j - \mathrm{Y}_i\right)}{R_{ij}^2} \tag{2.23}$$

The diagonal super-element $\mathbf{H}_{ii}$ involves all neighbors connected to site *i* and hence the second derivatives take the form

$$\frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_\mathrm{i}^2} = \gamma \sum_i \frac{\left(\mathrm{X}_j - \mathrm{X}_i\right)^2}{R_{ij}^2} \tag{2.24}$$

$$\frac{\partial^2 V_{ANM}}{\partial \mathrm{X}_i \partial \mathrm{Y}_j} = \gamma \sum_i \frac{\left(\mathrm{X}_j - \mathrm{X}_i\right)\left(\mathrm{Y}_j - \mathrm{Y}_i\right)}{R_{ij}^2} \tag{2.25}$$

Inversion of **H** yields 3*N*-6 non-zero normal modes, the frequencies and shapes of which are defined by the non-zero eigenvalues and corresponding eigenvectors of **H**. The k^{th} eigenvector, for example, is a 3*N*-*d* array of *N* blocks, each (of size 3x1) representing the fluctuation vectors of an individual residues in the k^{th} mode.

2.3.1.3 Calculations of B factors

The experimental Debye-Waller factors are related to the ms fluctuations of individual residues ($1 \le i \le N$) as $B_i = (8\pi^2/3) \langle(\Delta \mathbf{R}_i)^2\rangle$, which, upon substitution from equation 2.10, can be expressed in terms of the diagonal elements of $\mathbf{\Gamma}^{-1}$ as

$$B_i = \left(8\pi^2 k_B T/\gamma\right)\left[\mathbf{\Gamma}^{-1}\right]_{ii} \tag{2.26}$$

The trace of the 3x3 diagonal block elements $\left[\mathbf{H}^{-1}\right]_{ii}$ corresponds to $\left[\mathbf{\Gamma}^{-1}\right]_{ii}$ of the GNM.

The force constant γ is the only adjustable parameter of the theory for a fixed cutoff distance. Its magnitude, usually of the order of 1 kcal/(mol Å^2), is found by rescaling the theoretically predicted average of the ms residue fluctuations to match the value indicated by the experimental B-factors. The force constant γ affects the absolute size of fluctuations only, but not the relative fluctuation amplitudes or the distribution (shape) of the fluctuations of the individual residues. The agreement observed between theory and experiments is satisfactory in that no residue specificity and non-linear effects are included in the theory, on the one hand, and, the experiment is subject to uncertainties on the other. The average correlation between GNM fluctuation predictions and the B-factors was determined as 0.64 for a data set of 113 proteins (Kundu et al., 2002).

2.3.2 Structures in the dark state

The GNM and ANM were applied to two dark-state crystal structures of rhodopsin deposited in the PDB, 1L9H (Okada et al., 2002) and 1U19A (Okada et al., 2004), with 2.6 Å and 2.2 Å resolutions, respectively. Essentially identical slowest modes were found for these structures, consistent with the well-known robustness of such cooperative modes. Precisely, the first (slowest) GNM modes evaluated for 1L9H and 1U19a are verified to be almost indistinguishable, and equivalent to the linear combination of the first two modes (respective eigenvalues 0.29 and 0.32) computed by ANM for 1UL9A. The correlation cosine between the square displacements induced by GNM and ANM equivalent modes is 0.837. This mode was also verified to be the first (slowest) mode of motion for 1L9Ha. Based on these results, we concluded that this is the most robust mechanism of motion, and we reported (Isin et al., 2006) the results based on this mode, referred to as the 'global mode'. The succeeding (lowest frequency) mode of motion was observed to give rise to a bending in the overall structure, which would be strongly hindered by the presence of surrounding lipid molecules, if the latter were explicitly included in the model. A drop-off has been observed in inverse eigenvalue in the case of higher modes, which justifies their omission from further analysis.

2.3.3 Generation and refinement of ANM structures

We generated two sets of fluctuating conformations, referred to as 'positive' or 'negative' deformations. These conformations are represented by the $3N$-dimensional arrays of instantaneous position vectors and generated by rescaling the deformation induced by the global mode by the parameter s (Xu et al., 2003). We name these positive and negative deformations as

$\mathbf{R}^+(s)$ and $\mathbf{R}^-(s)$ separately. To determine which of these fluctuating conformations is more stable, and to correct for possible unrealistic distortions in bond lengths and bond angles, both sets of conformations were subjected to a short energy minimization scheme using MOE (Molecular Operating Environment) package, version 2001.01 (Chemical Computing Group, Inc., Montreal), with the Amber94 force field (Cornell et al., 1995). $\mathbf{R}^+(s)$ was observed to reach a substantially lower energy state compared to the $\mathbf{R}^-(s)$ (while both were higher than the energy minimized form of the dark state structure, as required), and the time elapsed to attain the equilibrated $\mathbf{R}^+(s)$ state was significantly shorter. $\mathbf{R}^+(s)$ was accordingly adopted as the Meta II model, using scaling factor, $s = 3$. It is important to remember that this scaling factor is arbitrary. The motions predicted by our model provide a qualitative understanding of motions during activation, They do not allow for predicting their absolute scales. However, the relative displacements found by the GNM/ANM are consistent with available experimental data (see Results and Discussion).

2.4 RESULTS

2.4.1 A hinge region that closely interacts with the β-ionone ring of the chromophore coordinates the collective dynamics of the protein

The global mode profile, i.e. the residue displacements $[\Delta\mathbf{R}_i]_1$ in the most cooperative mode of motion, predicted for rhodopsin in the dark state are presented as a function of residue index i in Figure 2.2A. The profile shows which regions move in the positive and negative directions along the principal axis of deformation induced by GNM mode 1, thus revealing which pairs of

structural elements are subject to correlated and anticorrelated fluctuations. Also crossover residues between anticorrelated regions are found from the intersection with zero line (dashed). Panel B in Figure 2.2 displays the corresponding normalized square amplitudes $[(\Delta\mathbf{R}_i)^2]_1$, also known as the *mode shape*, obtained with GNM (blue) and ANM (black), and panel C compares the theoretically predicted (red, GNM; blue, ANM) B-factors, $B_i = (8\pi^2/3)\langle(\Delta\mathbf{R}_i)^2\rangle$, and their experimental (X-ray crystallographic; black) counterparts. Note that the relative displacements of the structural elements are more distinctive upon examination of the global mode (panel B), as opposed to the B-factors that contain the contributions from a multitude of modes (panel C). Panel D maps the panel B onto a color-coded ribbon diagram. The most constrained regions (minima in the global mode) are shown in blue, and maxima in orange.

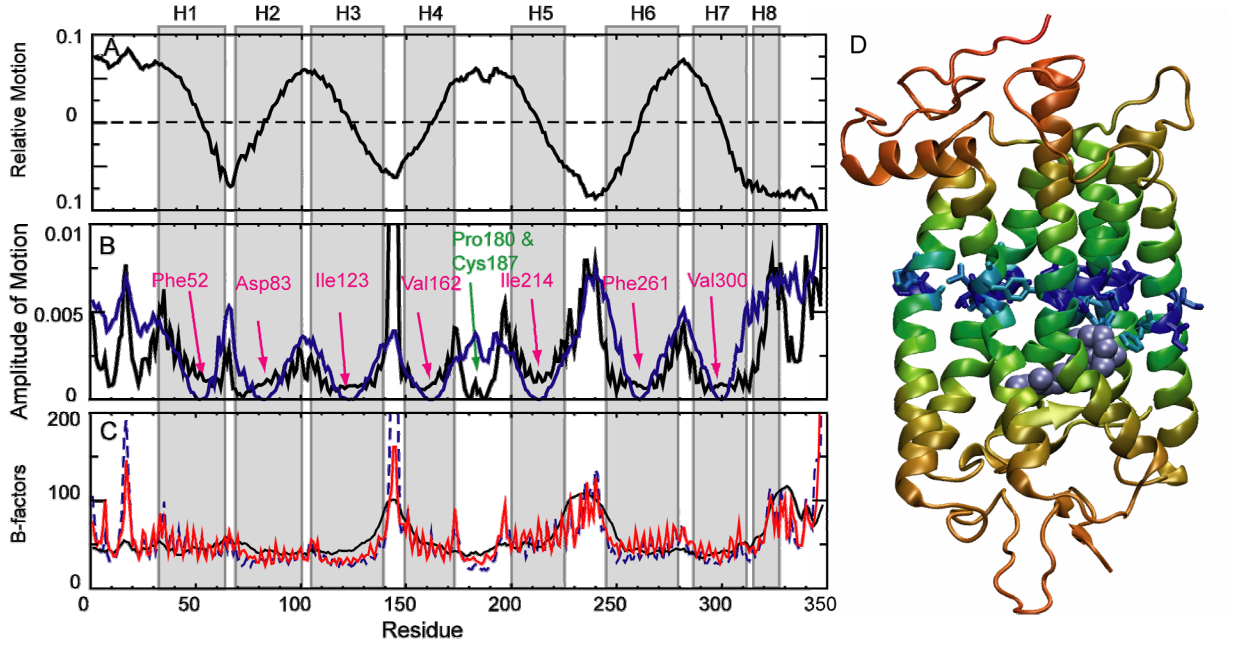

Figure 2.2 Mobilities and motions predicted by GNM and ANM
(A) The global mode profile calculated with GNM, indicating the relative motions of different regions of the proteins along the principal (normal) mode coordinate. The curve is directly given by the elements of the eigenvector associated with the slowest non-zero mode. Positive and negative regions delineate structural blocks subject to concerted motions. The zero-crossing points serve as hinges between anticorrelated regions. The locations of the helices (H1-H8) are indicated on the upper abscissa and distinguished by gray bands. (B) The distribution of

square displacements of residues predicted by the GNM (blue) and ANM (black). The non-TM regions exhibit higher mobilities in general, especially CL2 (between H3 and H4) and CL3 (between H5 and H6). Residues acting as hinge centers (minima near the middle of each TM helix) are labeled. The ANM yields two additional minima: Pro180 and Cys187 near the EC entrance to the chromophore binding pocket. (C) Experimental (black) and predicted thermal B-factors from GNM (red) and ANM (dashed blue). (D) Ribbon diagram of rhodopsin) color-coded according to the relative motions in panel B in the order of increasing mobility blue (lowest mobility), cyan, green, yellow, orange, red (highest mobility). The side chains are shown for the seven GNM hinges labeled in panel B and 11-*cis*-retinal is shown in light blue space-filling representation. The hinge site divides the protein into two anti-correlated regions: one on the CP side and the other containing the chromophore binding pocket on the extracelluar side. This image was generated using VMD (Humphrey et al., 1996).

The minima in the GNM/ANM global mode shapes usually point to residues implicated in mechanical roles relevant to biological function. Seven minima (Table 2.1, column 1) are identified here by the GNM, the centers of which are labeled in Figure 2.2B. Mapping these residues onto the ribbon diagram (Figure 2.2D) reveals that they all lie in the middle of the TM helices, about halfway between the EC and CP regions, thus dividing the protein into two halves that experience anticorrelated motions. We note that Gly121, Glu122, Leu125 and Phe212 among these residues directly interact with the β-ionone ring of the chromophore in the dark state. The rest of the chromophore binding pocket lies immediately below the hinge region. This suggests that the isomerization of 11-*cis*-retinal into all-*trans* form during activation (see chromophore binding site analysis, below) is a cooperative process that can engage the entire TM helices by interacting with the global hinge region. The hinge residues are poised to transmit the conformational changes that trigger the new, activated conformation from the retinal-binding pocket to the CP and EC ends of the molecule.

Largest amplitude motions are predicted for the loop regions. The CP loops tend to have larger mobilities than the EC loops (Figure 2.2B). Interestingly, the loop EC2 is much less mobile than the shorter CP loops CP2 and CP3. This suggests there is a possibly larger conformational motion upon retinal isomerization on the CP side than the EC side.

Table 2.1 Residues indicated by GNM/ANM to play a critical role in functional dynamics

Global hinge sites (minima in global mode shapes)	Amino acids affected by retinal isomerization			Peaks in high frequency modes(Rader et al., 2004)
	Stabilizing retinal in the dark state	Clashing with *trans* retinal in the dark state	Stabilizing all-*trans* retinal in the Meta II state	
Gly51, Phe52, *Pro53* (H1); Ala82, *Asp83*, Leu84 (H2); *Gly121*, *Glu122, Ile123*, Ala124, *Leu125*(H3); Trp161, Val162, Met163, *Ala164* (H4); Phe212, Ile213, Ile214 (H5); Ala260, *Phe261* (H6); Ala299, Val300, Tyr301(H7)	*Glu113(*H3) *Ala117*(H3) *Thr118*(H3) *Glu122*(H3) *Met207*(H5) *His211* (H5) Phe212(H5) *Phe261*(H6) *Trp265*(H6) Tyr268(H6) Ala269(H6)	*Thr118* (H3) *Cys167* (H4) Ala168 (H4) *Pro171* (H4) *Tyr178* (H5) *Gly188* (β4) Ile189 (β4) *Phe203* (H6)	*Thr118* (H3) *Cys167* (H4) Ala168 (H4) Ala169 (H4) *Pro171* (H4) *Leu172* (H4) *Tyr178* (β3) *Ser186* (β4) *Cys187* (β4) Ile189 (β4) *Phe203* (H5) *Trp265* (H6) Tyr268 (H6) *Ala295* (H7)	*Pro23 Asp83* Met86 *Val87* Phe91 *Thr94 Ser98* Gly101 *Tyr102* *Phe103 Cys110* *Glu113 Gly114* Phe115 *Ala117* Gly120 Ala124 *Tyr178 Pro180* *Glu181 Met183* *Gln184* Cys185 *Ser186 Cys187* *Ala292* Ala299

2.4.2 Opening of the helical bundle by a cooperative torsion of the transmembrane helices

The GNM analysis also provides information on the *size* of collective motions. Information on the *directions* of motions, on the other hand, is obtained by ANM analysis (see section 2.3.3). The global mode shape obtained by the ANM, shown by a black curve in Figure 2.2B, matches well its GNM counterpart. An additional minimum appears in the mode shape by ANM in loop EC3 containing the residues Pro180 and Cys187 which constrain the EC entrance to the chromophore binding pocket.

The structure deformed according to ANM (and refined by energy minimization (see Methods) is the predicted Meta II state. The differences between the dark state and the ANM state are shown in Figures 2.3 and 2.4. Figure 2.3 shows a side view of the two structures and Figure 2.4 presents a series (panels (A-E)) of cross-sectional slices separated by 10 Å along the

cylindrical axis, starting from the CP region (panel A). For reference, the residue numbers at H6 that are part of a given slice are also given. The comparison between the two structures reveals that there is an overall torsion or twisting of the molecule involving all of the TM helices. This is consistent with the identification of global hinges spread across all the TM helices (Figure 2.2B). The CP ends of all helices undergo a counterclockwise rotation in the ANM-predicted state, relative to their positions in dark state, when viewed from the CP end (panels A and B), while the opposite (EC) end of the helices display a clockwise rotation (panels D and E). The retinal binding pocket is located in panel D. Accompanying this torsional motion is an overall expansion in the cross-sectional areas at two ends of the TM region, while the central region remains almost unchanged.

There are a number of additional features predicted by ANM that merit a more detailed examination. First, H6 (magenta) rotates and extends into the cytoplasm as can be viewed in Figures 2.3B and 2.4A. Second, H3 (yellow) seems to assume a pivotal, central position (Figure 2.4C). In the ANM predicted structure, H3 is more exposed (than in the dark state) to the cytoplasm; In the dark state, it closely interacts with H4 (green) near the CP interface while simultaneously interacting with H2 (orange) near the EC interface (Figure 2.3A). Facilitating the exposure of its CP end to the cytoplasm is the motion of H4 away from H3, stretching the loop CL2, and creating a cavity that provides access to amino acids at H3 and CL2. This outward radial motion of the CP end of H4 is compensated by a motion in the opposite direction at its EC end, moving it closer to the retinal (see chromophore binding pocket analysis, below). Other helices also readjust their positions, resulting in an overall opening of the helical bundle at the CP side. H1 (red) forms a kink at Gly51-Pro53, resulting in a bending at the CP part of H1 away from the bundle. Similarly, H7 bends at Ala299 and also rotates around this residue. The soluble

and highly mobile helix H8 (cyan) rotates within the plane perpendicular to the TM helices, as can be seen in Figure 2.4A. At the EC side, the torsions of helices also result in an opening of the helical bundle, but of a smaller magnitude than at the CP side. The innermost pair of β strands, β3 and β4, is highly constrained and remains almost fixed in space (Figure 2.4E).

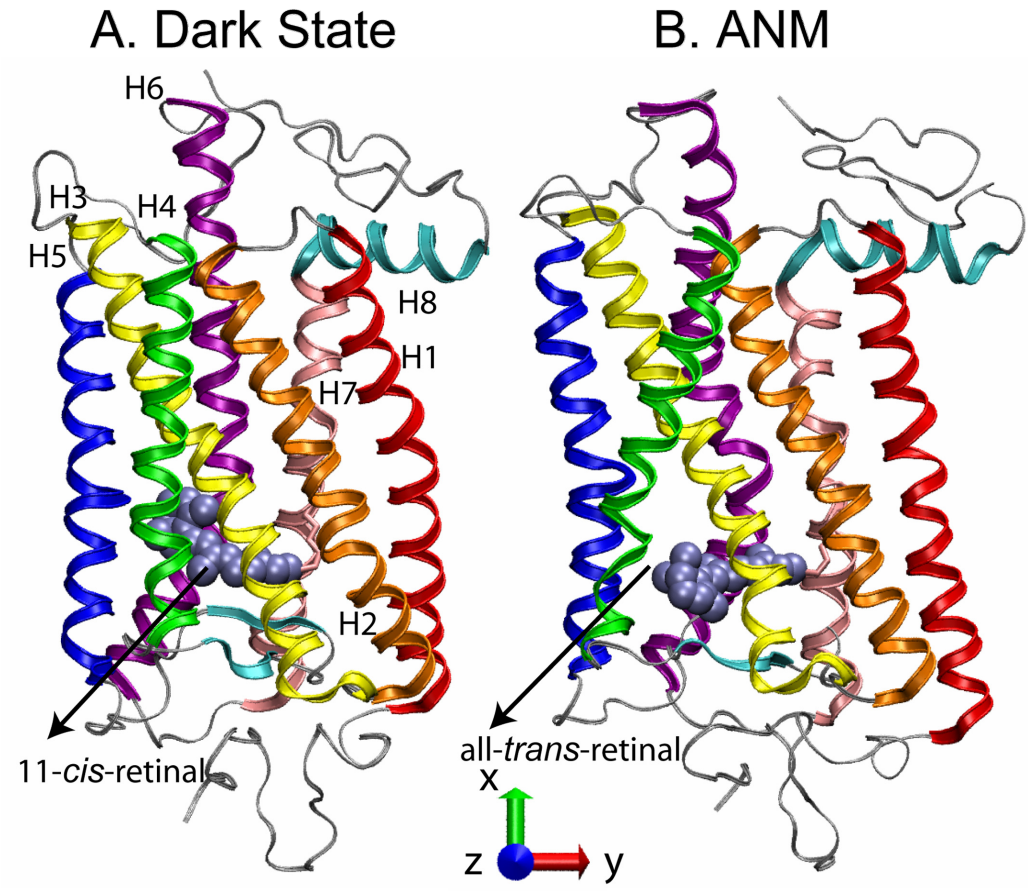

Figure 2.3 Comparison of the dark state (A) and ANM-predicted structure proposed for the Meta II (B) The dark and ANM structures contain the chromophore in the 11-*cis* retinal, and all-*trans* isomeric forms, respectively. This image was generated using VMD (Humphrey et al., 1996).

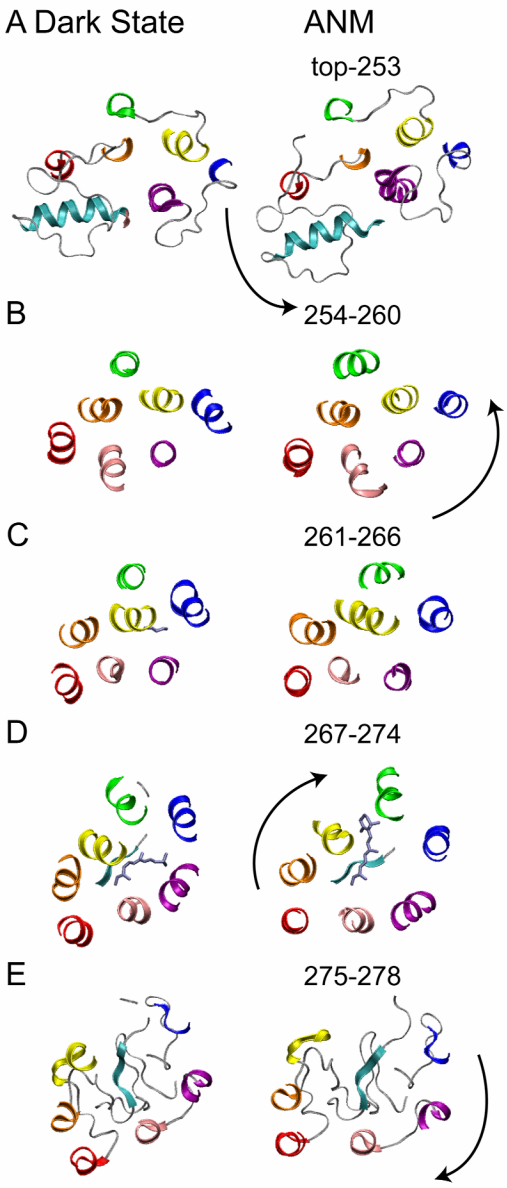

Figure 2.4 Comparison of the two structures viewed from top
Cross-sectional views of slices of 10 Å, displayed for dark state structure (left column) and ANM-predicted Meta II structure (right column). The slice corresponding to the chromophore binding site is shown in panel D. Note the counter-rotation of the CP and EC ends of the structure in the ANM structure, indicated by the arrows to guide the eye. Residue indices corresponding to H6 residues are given to indicate the position of the respective slices. This image was generated using VMD.

2.4.3 Chromophore binding pocket analysis: Dark-state structure cannot accommodate all-*trans*-retinal

The reconfiguration of 11-*cis* retinal in the dark state into all-*trans* form in the ANM structure can be seen in Figure 3 panels A and B, respectively. In order to understand more closely how all-*trans*-retinal triggers the destabilization of the dark state upon isomerization, the dihedral angle of the bond between its carbon atoms C_{11} and C_{12} was rotated by 180° within the context of the dark state structure, followed by a short energy minimization protocol that retains the retinal in the *trans* form. Figure 2.5 panels A and B show the neighborhood of the chromophore, in the *cis* and *trans* forms of retinal, respectively, when that the dark state structure remains unchanged. Opsin heavy atoms within 4.5Å of the retinal are displayed in each case. Because all-*trans*-retinal is more extended than 11-*cis*-retinal, isomerization leads to steric clashes with the opsin, unless relieved by suitable conformational rearrangements. The rhodopsin residues that show steric clashes with all-*trans*-retinal atoms are listed in Table 2.1, column 3. Most of these clashes involve the β-ionone ring of the chromophore. The large number of steric clashes confirms previous notions (Saam et al., 2002;Ishiguro et al., 2003) that a rigid dark state structure cannot accommodate the all-*trans*-retinal. Therefore, the molecule must adopt a new conformation with a larger chromophore binding pocket that can accommodate the all-*trans* isomer..

2.4.4 Photoisomerization changes the network of interactions in the chromophore binding pocket

The chromophore binding pockets in the dark state and in the ANM Meta II structure, are shown in Figure 2.5 panels A and C, respectively.Among the interactions between the opsin protein and the retinal, those involving the β-ionone ring are the most strongly affected by isomerization. This is due to the relocation of the β-ionone ring to a different environment upon isomerization and the rearrangement of the helices at the chromophore binding site to accommodate the new conformation of the retinal. In the dark state, the β-ionone ring is highly constrained by contacts made with H3, H5 and H6, (Figure 2.5A and Table 2.1 column 2). The ANM Meta II structure with all-*trans*-retinal, in contrast, identifies residues from H4, H3 and β-hairpin at the EC loop EL2 that stabilize the β-ionone ring (Figure 2.5C and Table 2.1 column 4). The most pronounced new constraints imposed on the ring are by the β-hairpin, in particular Tyr178 and Ile189, and by H4 such as Cys167, Ala168, Ala169, Pro171, and Leu172.

2.4.5 Comparison with structural changes experimentally observed upon activation of rhodopsin and other GPCRs

The current model for the mechanism of activation by light in rhodopsin and by ligands in GPCRs in general is essentially composed of two aspects; (1) intrinsic conformational changes, triggered by retinal isomerization which simultaneously (2) release structural constraints. Much work has focused on the identification of both the nature of conformational changes and the key structural constraints ("*microdomains*") shared by the GPCR family (Nakayama and Khorana, 1990;Sakmar et al., 1991;Nakayama and Khorana, 1991;Ridge et al., 1992;Cohen et al.,

1993;Samama et al., 1993;Hogger et al., 1995;Scheer et al., 1996;Ballesteros et al., 1998;Baranski et al., 1999;Rasmussen et al., 1999;Borhan et al., 2000;Alewijnse et al., 2000;Ballesteros et al., 2001;Shi et al., 2002;Visiers et al., 2002;Greasley et al., 2002;Klein-Seetharaman et al., 2004). The microdomains include (i) an ionic interaction region between the charges of ligand and receptor, corresponding to retinal Schiff base and the Glu113 counter-ion in rhodopsin (Sakmar et al., 1991;Cohen et al., 1993) (ii) a D(E)RY motif at the CP end of H3 and the $X_1BBX_2X_3B$ motif at the CP end of H6 (B, basic; X, non-basic) (Samama et al., 1993;Hogger et al., 1995;Scheer et al., 1996;Ballesteros et al., 1998;Rasmussen et al., 1999;Alewijnse et al., 2000;Ballesteros et al., 2001;Shi et al., 2002;Visiers et al., 2002;Greasley et al., 2002) (iii) an Asn-Asp pair at H1 and H2, respectively, and a NPXXY motif at H7(Okada et al., 2002) and (iv) an aromatic cluster surrounding the ligand binding pockets (Nakayama and Khorana, 1990;Nakayama and Khorana, 1991;Ridge et al., 1992;Borhan et al., 2000). In the following, we will compare the experimentally observed changes to those predicted by the ANM.

2.4.6 Changes observed in retinal-rhodopsin contacts upon activation

Photoaffinity labeling and cross-linking experiments indicate that retinal isomerization is accompanied by a change in the neighboring amino acids, in particular those contributing to the aromatic cluster (microdomain iv) that coordinate the chromophore. The β-ionone ring of 11-*cis*-retinal can be cross-linked to Trp265 at H6 in the dark state, but not after light-activation (Nakayama and Khorana, 1990). Instead, Ala169 from H4, a residue located >10 Å away from the ionone ring in the dark state (Palczewski et al., 2000), forms a cross-link with retinal in the light, but not in the dark (Borhan et al., 2000). These observations are in good agreement with

our findings, illustrated in Figure 2.5. In the dark state, Trp265 is located immediately adjacent to the β-ionone ring towards the CP side, while Ala169 does not interact with the retinal (Palczewski et al., 2000). In contrast, in the predicted active structure with all-*trans*-retinal, Ala169 interacts closely with the atoms of β-ionone ring while Trp265 does not.

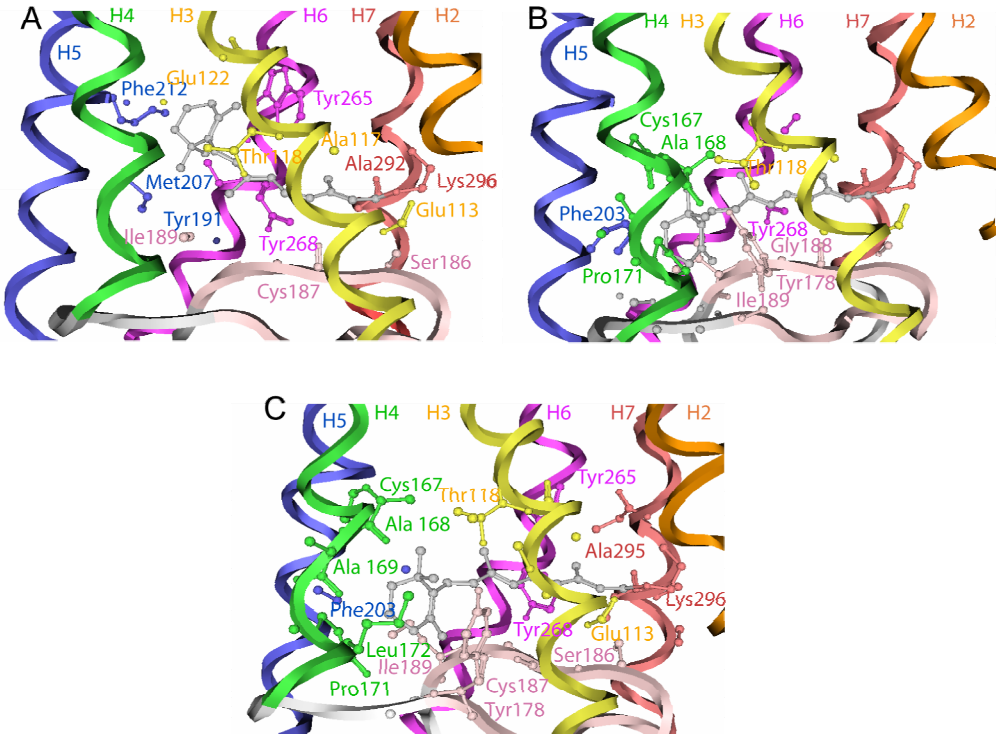

Figure 2.5 Close neighborhood of the chromophore in the dark state before (A), and after (B) isomerization to *trans*-form, and in the ANM predicted form (C).

(A) The dark state structure contains retinal in the 11-*cis* form. Its β-ionone ring is stabilized by Glu122 at H3, Met207 at H5, His211 at H5, Phe212 at H5, Phe261 at H6, Trp265 at H6, Ala269 at H6. (B) There are steric clashes between all-*trans*-retinal and opsin residues in the chromophore binding pocket in the dark state structure of rhodopsin. All heavy atoms of opsin within 4.5Å of all-*trans*-retinal are displayed. The residues that clash with all-*trans*-retinal are Thr118, Cys167, Ala168, Pro171, Tyr178, Gly188, Ile189 and Phe203. (C) The ANM Meta II structure with *all-trans* retinal where all steric clashes are relieved. The polyene carbons C9-C15, C19 and C20 have common neighboring residues in both structures. These residues include Thr118 at H3, Tyr268 at H6, Cys187 and Gly188 at β4, Gly114, Ala117 and Glu113 at H3, Lys296 at H7, and Ile189 at β4. In addition to these common residues, Tyr178 and Cys181 at β3 interact with the polyene chain of the retinal at the ANM Meta II structure. Tyr178 runs parallel to the β-ionone ring of all-*trans*-retinal. This image was generated using VMD (Humphrey et al., 1996).

2.4.7 Experimental evidence for the opening of the CP side of TM helical bundle

Extensive Cys scanning mutagenesis along the CP surface of rhodopsin in combination with site-directed spin labeling followed by EPR analysis of mobility, accessibility and spin-spin interactions, sulfhydryl reactivity, and disulfide cross-linking rates measurements have been performed to deduce the molecular nature of the conformational changes accompanying activation (Farahbakhsh et al., 1995;Farrens et al., 1996;Altenbach et al., 1996;Cai et al., 1997;Altenbach et al., 1999a;Altenbach et al., 1999b;Meng and Bourne, 2001;Klein-Seetharaman et al., 2001;Hubbell et al., 2003). In particular, the mobility of spin-labeled side chains at the buried surfaces of H1, H2, H3, H6, and H7 are found to increase upon isomerization (Hubbell et al., 2003). This indicates reduced packing in the core of the protein. Furthermore, several sets of pairs of Cys suggest that there is an increase in the relative distances between the CP ends of helices, especially between H3 and H6 (Hubbell et al., 2003). Six of these disulfide bridges (138-251, 139-247, 139-248, 139-249, 139-250, 139-251), connecting H3 and H6 close to their CP end, prevented the light activation of rhodopsin (Farrens et al., 1996;Cai et al., 1999). This indicates that the relative positions of H3 and H6 change during activation. Figure 2.6A illustrates the relative positions of these disulfide-bridge forming sites in the dark state (transparent) and in the Meta II ANM structure (opaque). A smaller displacement in the CP ends of H1 and H2 away from H7 has also been detected experimentally (microdomain iii), consistent with the ANM-predicted increase in surface accessibility at the CP region. It is noteworthy to observe that the surface accessibility of the G-protein contact site near the ERY motif (microdomain ii) increases upon transition from the dark state structure to the ANM-predicted deformed structure. The surface accessibility of the ERY motif residues Glu134 and Arg135 in particular increases from 23.6 $Å^2$ to 29.7 $Å^2$ and from 14.2 $Å^2$ to 25.5 $Å^2$ respectively,

while Tyr136, being already solvent-exposed, does not exhibit a significant change in solvent accessibility. Figure 2.6B shows the ERY motif both in the dark state (transparent) and ANM-predicted deformed state (opaque) and change in the relative positions of the CP ends H3, H4 and H6.

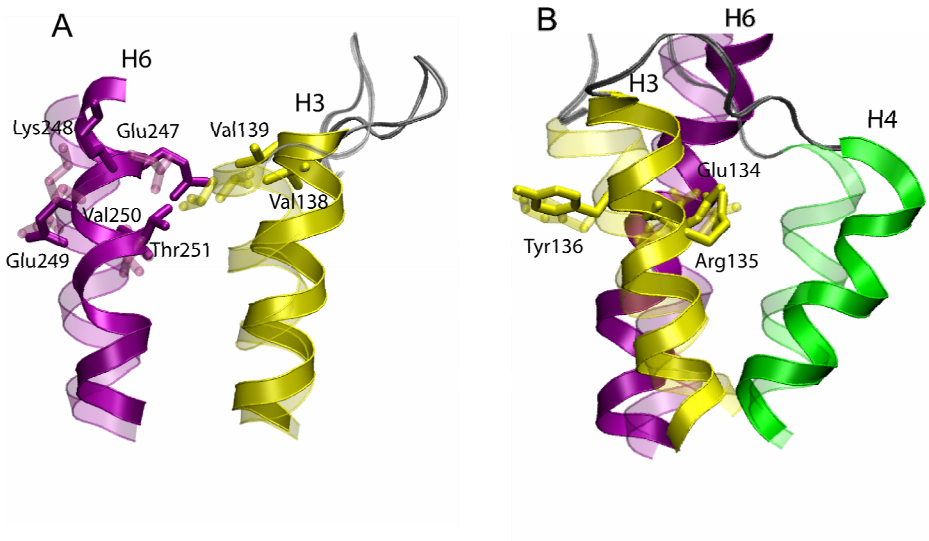

Figure 2.6 Repositioning of CP ends of H3, H4 and H6 in the ANM Meta II (opaque) and dark state (transparent) structures.
(A) Relative movements of H3 and H6. The residues shown as sticks are detected by disulfide cross-linking studies to inhibit the light activation of rhodopsin when mutated to Cys (Farrens et al., 1996). (B) Position of the ERY motif in the dark state and the ANM Meta II structures, showing how the ERY motif becomes more surface-exposed in the proposed Meta II form due to the local rearrangements of helices in its close neighborhood, including in particular the displacement of H4 away from H3 and H6. The surface accessibility of the ERY motif residues Glu134 and Arg135 increases from 23.6 $Å^2$ to 29.7 $Å^2$ and from 14.2 $Å^2$ to 25.5 $Å^2$ respectively, while Tyr136 being already solvent-exposed does not exhibit a significant change in solvent accessibility. This image was generated using VMD (Humphrey et al., 1996).

The above comparison with experimental observations support the ANM predicted structural changes, including those affecting the microdomain constraints. In the next section, we will discuss the new insights gained by GNM/ANM analysis on the activation mechanism. This includes an explanation of the effects of mutations on Meta II stability, how refinements using atomic details impact the model for rhodopsin activation, and a mechanistic explanation of the observed conformational changes and their implications for G protein binding and activation.

2.4.8 Metarhodopsin II fluorescence decay measurements

The stability of the light-activated state can be quantified by measurement of the decay rate of Meta II, assuming that the destabilization of the equilibrium state will induce faster decays. To investigate position-specific contributions to Meta II stability, we collected the effects of 228 site-directed amino acid replacements in rhodopsin on the rates of Meta II decay from a survey of published and experiments conducted in the laboratory of Dr. Klein-Seetheraman for this particular study. 97 of these corresponded to unique positions in the rhodopsin sequence. In all cases, the Meta II decay was determined using the fluorescence assay developed by Farrens and Khorana (Farrens and Khorana, 1995). Each mutant was measured using wild-type as a control. In order to compare different studies, we divided the mutant decay rates by the corresponding wild-type decay rates for a given set of experiments. A mutation was considered to have an effect on Meta II stability, if the rates of decay were either enhanced or decreased by at least 20% with respect to the wild type (WT) value. If a residue was predicted to be important (by GNM/ANM), but a Meta II decay rate had not been measured, we used data from misfolding experiments when available, since mutants that are not stable in the dark are not expected to be stable in the light either. When there were multiple experiments including different mutation

backgrounds and different substitutions on the positions in question, we adopted the data supported by the larger majority of experiments. For eight residues, the experimental results were contradictory; therefore, we did not include those in our analysis. In addition, we ignored the effects observed by drastic mutations, such as Gly to Glu, which would be expected to affect stability irrespective of a role in activation or stability. Thus, we generated a non-redundant dataset containing data on Meta II decay or dark state stability for 119 unique residues.

2.4.9 Analysis of Meta II decay rates of rhodopsin mutants in light of key residues identified by GNM/ANM

Empirical characterization of rhodopsin mutants typically includes investigation of Meta II stability by way of quantifying Meta II decay rates (Farrens and Khorana, 1995). While a useful measure to estimate the role of a given amino acid in structure and function, Meta II decay rates have been mostly phenomenological, providing little insight into the mechanisms by which stability is altered. Here, we provide such mechanistic insights using GNM/ANM results.

We presently identified three groups of residues of interest, listed in Table 2.1: (i) those participating in the global hinge region, (ii) those directly affected by retinal isomerization (Figures 2.4 and 2.5) and (iii) those emerging as peaks in the GNM/ANM fast modes, i.e. distinguished the high frequency modes. The former two groups are identified in the present study. Details of the GNM/ANM fast modes will be given in the following chapter where its results combined with First analysis to find folding cores and stable regions of the protein

(Rader et al., 2004) will be presented. These three groups of residues are colored red, blue and green,

respectively, in Figure 2.7 panel A. Note that some of the residues are included in more than one group, overall summing to 61 residues implicated in functional interactions/dynamics (Table 2.1).

For comparative purposes, we compiled a comprehensive list of Meta II decay rates in mutants of rhodopsin from the literature (the extensive list of Meta II decay rate experiments is available at http://www.ccbb.pitt.edu/People/isin/MetaIIdecaytable/) and further experiments conducted Dr. Klein-Seetheraman's labs for this study (see below and Table 2.2). In the absence of Meta II decay rates, we resorted to data from misfolding experiments when available. This led to a dataset of 127 distinctive residues with known Meta II or dark state stability. If the ratio μ of the mutant decay rate to WT decay rate ($\mu = r_m/r_{wt}$) lies within the range $0.8 \leq \mu \leq 1.2$, the mutation is assumed to be inconsequential; otherwise, the mutation is assumed to affect stability/function. Based on this criterion, 45 amino acid replacements were found to have an effect on Meta II stability and/or misfolding, while 74 were inconsequential. 8 of the 127 residues had contradictory results and have not been included in our comparative analysis.

Table 2.2 Rhodopsin mutants and their Meta II half life.

Mutant	Meta II half life (min)	Ratio between Meta II half lives of mutants and the WT
WT	15.9	1.00
C140S	15.1	0.95
C167S	25.5	1.60
C222S	12.8	0.81
C264S	18.9	1.19

Our objective was to see if the 61 residues inferred by GNM/ANM to play a critical role (Table 2.1) were also observed by experiments to impair decay or folding properties (true positives, TPs), while the other residues (not predicted by GNM/ANM to play a key role) were confirmed to be inconsequential (true negatives, TNs). The comparison between theory and experiments yielded 38 TPs (written in italic in Table 2.1, colored green in Figure 2.7, panel B), 1 false positive, 73 TNs, and 7 false negatives. In 111 out of 119 experimentally examined cases GNM/ANM was therefore able to correctly predict the experimental outcome. Thus, the GNM/ANM results can help explain 93% of the experimental data on mutant's Meta II decay rates (and folding). We note that experimental data are not (yet) available for 21 residues (colored purple in Figure 2.6, panel B), and ambiguous for one residue, indicated to be important by GNM/ANM analysis. The observed validation with available experimental data, whenever available, strongly supports the use of GNM/ANM predictions for guiding future experiments. It remains to be seen if these 21 residues play an important for Meta II stability or folding properties. In particular we note among them a few aromatic residues, Tyr301, Trp161, and Phe212, which appear to play a central role in stabilizing the global hinge region (Figure 7B).

The accuracies of the above calculations are limited by the fact that the mutations reported in the literature vary in their severity. For example, replacement of a Cys by an Arg is more severe than replacement by a Ser, especially in a membrane environment. To provide further testing of the predictive value of GNM/ANM Meta II effect analysis, we generated a small dataset in which Cys residues at positions 140, 167, 222 and 264 were replaced with Ser one at a time and the Meta II decay rates were measured under identical conditions. Table 2.2 lists the results. The only significant effect was by the Cys167Ser substitution. The comparison of the dark and ANM structures provides a plausible explanation for this result. In the dark

structure, Cys167 shows steric clashes with C1, C2 and C16 atoms of all-*trans*-retinal. This result validates our approach, since the increase in Cys167Ser Meta II stability cannot be explained by inspection of the dark state crystal structure. Instead, it can be well understood from the differences in the retinal binding pockets in the dark and in the ANM structures. For the other three cysteines investigated experimentally here, ANM/GNM correctly predicted that amino acid replacement at these sites does not have an effect on Meta II stability.

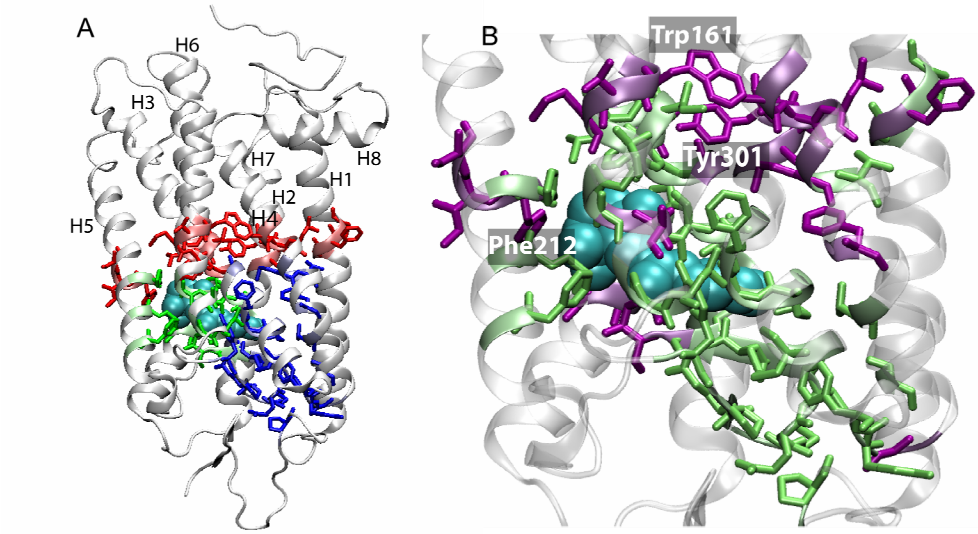

Figure 2.7 Residues identified by GNM/ANM to play a key role in activation
(A) Residues identified by GNM/ANM to play a key role colored by their different roles in the global mechanics and/or function; global hinge sites in red, amino acids affected by retinal isomerization in blue, and peaks in high frequency modes in green (B) A closer look of the same subset of residues identified by GNM to play a key role. Those confirmed by Meta II and folding experiments to play a critical role are colored green, and the rest, purple. Both structures refer to the dark state conformation with 11-*cis* retinal shown in space-filled model colored cyan. This image was generated using VMD (Humphrey et al., 1996).

2.4.10 Refining the current model for activation of rhodopsin

The predicted Meta II model identifies several previously unknown details. For example, H4 emerges as a structural element playing an important role in transmitting the activation signal from the chromophore binding pocket to the cytoplasm. Triggered by steric clashes within the chromophore binding region, H4 undergoes significant rigid-body reorientation with respect to the cylindrical axis, and this reorientation is reflected in large displacements at both the CP and EC ends. At the chromophore binding pocket the displacement of H4 is more than 4 Å, outward the helical bundle. In particular the CP end of H4 moves away from the helical bundle resulting in an extension of CL2 between H3 and H4 and exposure of the ERY motif at H3 (constraint ii, above; see Figures 2.3B and 2.6B). Thus, in contrast to the prevailing view of activation (Meng and Bourne, 2001), it is not simply the relative displacement of H3 and H6 with respect to each other that produces the critical conformational changes recognized by the G protein. Likewise, a decrease in the mobility of Thr70 and Tyr74 at H2, Arg147 and Phe148 in CL2 have been attributed to the movement of H2 towards H4 (Farahbakhsh et al., 1995;Hubbell et al., 2003). The ANM analysis shows that H4 moves closer to H2.

The motion of H6 deduced from increases in the accessibility of Val250 and Thr251 at H6 upon activation (Altenbach et al., 1996) and the changes in distances of a range of spin labels at H6 with respect to Val139Cys at H3 (Farrens et al., 1996) were previously attributed to the movement of H6 away from the helical bundle so that it interacts less with helices such as H3 and H7. The GNM shows that four CP turns of H6 are extremely mobile and tend to extend towards the cytoplasm facilitated by a screw-like torsional rotation of H6.

Finally, conformational changes in the EC domain have not been investigated extensively by experiments previously. Our ANM results show that the opening of the helical bundle in the CP domain through torsional rotation in one direction is coupled to an opposite direction rotation in the EC side, which also induces an opening of the helical bundle at the EC end. This observation is in accord with findings that the disulfide bond at the EC-TM interface becomes accessible to reducing reagents upon light activation. The results seem to support the hypothesis that EC2 moves upon activation, releasing constraints imposed by its interactions, with helices H4 and H5 that help lock the receptor in an inactive conformation in the dark (Klco et al., 2005;Massotte and Kieffer, 2005). Further quantitative testing the ANM predictions will require distance measurements between spin labels placed in the CP domain and EC domain simultaneously. Double electron-electron resonance (DEER) experiments will allow such long-range distance measurements.

2.4.11 A mechanistic explanation for experimentally observed conformational changes in rhodopsin and implications for the activation of G proteins

The transmission of the signal of retinal isomerization to the CP domain, and induced opening of the helical bundle CP end to allow for G protein binding is ensured by a global hinge site spatially located in the close neighborhood (and directly interacting with) the retinal-binding pocket (Figure 2.2B and D). This site divides rhodopsin into two coupled but anticorrelated regions: one on the G-protein binding, CP side and the other on the EC side surrounding the chromophore binding pocket. The global motion coordinated by the hinge region is essentially manifested by increasing mobility towards the CP portions of H3-H6. The interactions at the hinge site with the retinal, predominantly with its ionone ring, restrict H6. Isomerization of the

retinal into the *trans* form releases these interactions, resulting in an extension of H6 towards the cytoplasm and rotation around its own axis (Figures 2.3B and 2.4A). On the other hand, *cis-trans* isomerization potentially entails steric clashes between H4 and the β-ionone ring, which are relieved by the repositioning of H4. Furthermore, while H5 stabilizes 11-*cis*-retinal dark state opsin interactions together with H3 and H6, it bends at its hinge and therefore interacts less with all-*trans* retinal in the activated state (Figure 2.3B). This combination of release of constraints or weakening of interactions, accompanied by the formation of new interactions is readily achieved by an overall torsional motion of the molecule driven by the first (global) collective mode. Thus, there is a clear cooperativity between the changes in the chromophore environment and the conformational changes in the CP domain upon light-activation, which can be explained by the intrinsic dynamic properties of rhodopsin in the dark state.

In summary, our results support the view that the dark state structure of rhodopsin is predisposed to undergoing the large-scale structural changes that are experimentally observed to be triggered upon light-induced retinal isomerization. This predisposition is unambiguously revealed by the global vibrational or relaxational fluctuations naturally favored by the topology of inter-residue contacts before light-activation.

It is a generally accepted notion that the opening of the helical bundle is the critical event for binding and activation of G protein, transducin, by (1) generating a cavity that provides sufficient space/volume for interaction and (2) exposing specific residues involved in direct recognition and binding (Meng and Bourne, 2001). Our model suggests that beyond these two general notions, we can now state that an overall torsional motion of the TM domain, that simultaneously induces the opening of the CP end of the TM bundle and the screwing motion of

H6, underlie the conformational transitions that facilitate transducin binding leading to GDP/GTP exchange.

2.5 DISCUSSION

In the present chapter, we presented a robust model for the light-activated Meta II state of rhodopsin, the prototypical member of the GPCR family. We show that information for functional conformational changes as they occur upon formation of the activated structure of rhodopsin, Meta II, is encoded largely in its inactive dark state structure. The relevance of the modeling results to the conformational changes conducive to the Meta II state is validated by a large amount of experimental data. More importantly, the newly proposed Meta II state model reveals previously unknown molecular details. Only a subset of these newly identified motions has been studied experimentally so far, and the model therefore provides a useful framework for further experimental validation.

It appears that there are only a few specific contacts that restrict the structure in a "locked" dark-state, while the overall molecule possesses the intrinsic potential to undergo highly concerted fluctuations and efficiently transmit signals from the chromophore binding domain across the TM region to both CP and EC domains. The inference is that it is not necessary to break and form a multitude of specific contacts within nanoseconds after retinal isomerization in rhodopsin (or ligand binding in other GPCRs), but only a few near the central hinge region in order to induce allosteric changes in global conformation. While the case for the role of internal constraints release in the GPCR activation process is very strong, its precise

mechanism has not been clear to date. The results described in this paper provide the first evidence for the *relaxation dynamics* of rhodopsin naturally favored by the inter-residue contact topology in the dark state. They also confirm that only a few strong constraints exist, while the bulk of the molecule enjoys conformational mobility, as previously suggested (Klein-Seetharaman et al., 2002;Klein-Seetharaman et al., 2004).

By definition, the motions predicted by the GNM/ANM are those along the lowest ascent pathway away from the global minimum in a multidimensional free energy landscape. The proposed conformational changes are therefore the most favorable ones which the dark-state will tend to undergo once a structural change has been triggered. Interestingly, these conformational changes provide an optimal 3D setting for (i) relieving the intermolecular steric clashes that would otherwise occur upon *cis* to *trans* isomerization of the retinal, and (ii) expose functional residues at the CP region, that can readily bind substrates such as the G protein.

We note that the GNM/ANM method relies upon the assumption that our structure is native-like and motions are similar to that native minimum in some coarse-grained energy landscape. It is clear that rhodopsin activation can be effected by changing the lipid environment; in our analysis we make no claims about how such changes to the lipid concentration or composition would impact our calculations. By not explicitly including the lipid in our model, we assume that its overall concentration/composition changes little during the conformational changes. On the other hand, rhodopsin activation has been pointed out to depend on the nature of the surrounding lipids. See for example; the work of Brown and collaborators (1994) and the more recent work of Botelho et al (2002). Understanding the effect of lipid

environment on the *intrinsic dynamics* of the protein would be a topic of out utmost interest and it will be explored by the new methodology introduced in the following chapter. However, the present chapter focused on the identification of the intrinsic dynamics of rhodopsin, based on the premises that (i) the topology of inter-residue contacts essentially determines the global motions of proteins, and (ii) the global modes are robust, i.e. they are uniquely defined by the equilibrium structure and are insensitive to the details of atomic interactions, as elaborated in several recent coarse-grained analyses of biomolecular equilibrium dynamics (Cui and Bahar, 2006). A recent analysis of the global dynamics of a series of potassium channels lend further support to the utility of examining the global dynamics of the proteins and identifying functional mechanisms (e.g. gate opening), regardless of the perturbations induced in the collective dynamics by the lipid environment (Shrivastava and Bahar, 2006).

The predisposition of structures to undergo functional changes has been pointed out in a number of recent studies (Tama and Sanejouand, 2001;Xu et al., 2003;Bahar and Rader, 2005;Ma, 2005;Eisenmesser et al., 2005;Changeux and Edelstein, 2005;Tobi and Bahar, 2005). A noteworthy example is the transition of hemoglobin from tense (T) to relaxed (R2 – CO bound) form (Xu et al., 2003), which was explained by the dominance of entropic effects (relaxation towards the least constrained state near the original state) in driving the functional reconfiguration of the molecule. The global torsion of rhodopsin that releases the constraints near the chromophore binding pocket and induces a weakening of intramolecular contacts at CP and EC regions also seems to be driven by entropic drives defined by the overall architecture. Notably, the most critical region controlling the collective changes in structure is located in the immediate neighborhood of the chromophore binding pocket, which is also the center of

chemical activity/transition. Such a coupling between global mechanics and chemical activity demonstrated for a series of enzymes (Yang and Bahar, 2005) appears to be a general design principle for proteins.

3.0 IDENTIFICATION OF CORE AMINO ACIDS STABILIZING RHODOPSIN

3.1 ABSTRACT

In many cases, rhodopsin dysfunction has been linked to a point mutation that perturbs the native structure or causes the misfolding of the protein. Here, we identify the structural elements that stabilize rhodopsin by computational analysis of the rhodopsin structure and comparison with data from previous *in vitro* mutational studies. We simulate the thermal unfolding of rhodopsin by breaking the native-state hydrogen bonds sequentially in the order of their relative strength, using the FIRST software (Jacobs et al., 2001). Residues most stable under thermal denaturation are part of a *core* which is assumed to be important for the formation and stability of folded rhodopsin. This core includes the Cys110-Cys187 disulfide bond at the center of residues forming the interface between the TM and the EC domains near the retinal binding pocket. This highly conserved disulfide bridge has been shown to be critical for correct folding of rhodopsin. The GNM high frequency modes of rhodopsin are also analyzed to identify the disulfide bond and the retinal ligand binding pocket as the most rigid region. Experiments confirm that 90% of the residues predicted by the FIRST software as core residues, cause misfolding upon mutation.

3.2 INTRODUCTION

A large number of naturally occurring point mutations has been found in rhodopsin (Berson, 1993;Dryja and Berson, 1995). The majority of these mutations are associated with *Retinitis Pigmentosa* (RP), a disease that ultimately leads to photoreceptor degradation and vision loss. RP mutations are found in each of the three domains of the rhodopsin structure: CP, TM, and EC. However, most of the mutations reside in the TM and EC domains (Sung et al., 1991;Inglehearn et al., 1992;Berson, 1993;Macke et al., 1993;Dryja and Berson, 1995). Mutations in the EC domain cause partial or complete misfolding, where misfolding is defined as the loss of ability to bind 11-*cis*-retinal. Studies of both naturally occurring RP mutations and designed mutations in the EC domain showed that misfolding was associated with the formation of a non-native disulfide bond, i.e. between Cys185 and Cys187 instead of Cys110-Cys187. The evidence for this came from folding studies of a rhodopsin mutant (Cys110Ala) and of the RP-inducing mutants: Cys110Phe and Cys110Thr (Hwa et al., 1999). Another study identified the formation of this incorrect, Cys185–Cys187, disulfide bond after reconstitution of rhodopsin from two complementary fragments (Kono et al., 1998). Further, studies of RP mutations in the TM domain of rhodopsin showed that they also cause misfolding by formation of an abnormal EC disulfide bond (Garriga et al., 1996;Hwa et al., 1997), which has been identified by mass spectrometric analysis to be between Cys185 and Cys187 (Hwa et al., 2001). Identification of the abnormal disulfide bond to be the same regardless of whether the mutation was in the TM or EC domain suggests that packing of the helices in the TM domain and folding to a tertiary structure in the EC domain are coupled (Khorana, 2000).

We have simulated the unfolding of rhodopsin using the FIRST software (Jacobs et al., 2001) and analyzed its stability in terms of the group of residues that form the core of structural stability. Residues that play a key role in maintaining the folded state have been further examined by generating high frequency modes of GNM (Demirel et al., 1998). These two computational methods generate predictions about the potential flexibility or mobility of each protein residue, respectively. Both have been applied to predict residues that are crucial for folding and stability (Demirel et al., 1998;Bahar et al., 1998). They were applied here to the crystal structure of rhodopsin (PDB code 1L9HA) (Okada and Palczewski, 2001) and compared to known experimental mutation and deletion data to assess the residues responsible for folding. The predicted core includes amino acids at the interface between the TM and EC domains and provides molecular explanation for the experimentally observed importance of the structural coupling between the EC and TM domains for the folding and stability of rhodopsin.

3.3 METHOD: FIRST

The FIRST software uses techniques from graph theory, to analyze and quantify the rigidity or flexibility of each bond in the protein (Jacobs et al., 2001). Each bond is identified as either flexible (free to rotate) or rigid (non-rotatable), according to geometric criteria (Hespenheide et al., 2002;Rader et al., 2002) using the atomic positions of an input PDB file (Berman et al., 2000). For rhodopsin, all-atom calculations were performed on molecule “A” from PDB entry 1l9h along with all ligands and buried water molecules. Buried waters were defined by ProAct (Williams et al., 1994) and polar hydrogen positions were optimized for hydrogen bonding by WHATIF (Vriend, 1990). Groups of atoms connected to each other via rigid bonds form rigid

clusters. The location of the rigid and flexible bonds can be mapped onto the 3D structure for visualization, using suitably colored ribbon diagrams. Alternatively, the rigid clusters can be schematically shown in a reduced (one-dimensional) representation of the protein sequence (Figure 3.1). Each line in the figure shows the backbone rigid cluster decomposition mapped onto the protein sequence corresponding to a different stage of unfolding. The main-chain bond is thus displayed as a thin black line if it is rotatable or as a colored block if it is part of a rigid cluster. This reduced representation allows for easier visualization of the denaturation profile, i.e. time evolution of rigid/flexible clusters during simulated unfolding. The FIRST software is accessible at http://firstweb.asu.edu.

3.4 RESULTS

3.4.1 Simulated thermal unfolding

The FIRST software was used to calculate the rigid and flexible regions in rhodopsin by modeling the structure as a network of rotatable bonds defined by the covalent bonds, hydrogen bonds, salt bridges and hydrophobic interactions (Williams et al., 1994;Jacobs et al., 2001). Each bond is identified as either flexible (free to rotate) or rigid (non-rotatable), according to graph theory considerations (Rader et al., 2002). Groups of atoms connected to each other via rigid bonds form rigid clusters. Figure 3.1 shows the so-called *dilution profile* (Hespenheide et al., 2002) obtained by simulated thermal denaturation of rhodopsin (Rader et al., 2004). Each line gives a schematic representation of the distribution of the rigid (thick, colored) and flexible (thin, black) regions in the molecule as the structure gradually unfolds. Each color represents a

different rigid cluster, which can include sequentially distant but spatially close segments because the rigid regions depend upon the underlying 3D structure. Thermal denaturation is simulated by breaking the hydrogen bonds of the protein one at a time in order of their strength from weakest to strongest (Hespenheide et al., 2002;Rader et al., 2002). Often hydrogen bonds are removed from rigid regions without causing any change in the rigid or flexible regions. These inconsequential steps are omitted from the dilution profile in Figure 3.1 so that only steps yielding different rigid region decompositions from the previous lines are shown.

For a given dilution plot, the transition state is defined by the peak in the second derivative of the number of "floppy modes" with respect to the mean coordination, <r>. These floppy modes quantify the available degrees of freedom in the protein. A previous study found a universal unfolding transition for proteins at $\langle r \rangle = 2.40 \pm 0.05$ (Rader et al., 2002). Additionally, the folding core is defined as the last (lowest) line in the simulated denaturation where at least three residues of two or more secondary structures are part of the same rigid cluster (color). This definition has been shown to correlate well with slow exchanging (folding) cores indicated by hydrogen/deuterium exchange experiments of proteins (Hespenheide et al., 2002;Rader et al., 2002;Bahar and Rader, 2005).

Scanning down the dilution profile in Figure 3.1, the reduced amount of red-colored regions corresponds to a loss in rigidity as hydrogen bonds are removed. Labels A through C along the left side of the figure indicate specific points along the unfolding pathway. Even in the initial state (first line of the denaturation profile) there are flexible regions, i.e. the loop CL2 and adjoining NT of H6, and the CT shown as thin black lines, an observation supported by previous studies on the dynamic properties of the CP domain including the CP ends of helices, using a

combination of cysteine mutagenesis and EPR spectroscopy (Columbus and Hubbell, 2002) and NMR spectroscopy (Klein-Seetharaman and Reeves, 2002). Between this mostly rigid state and the native-like state A, the loss of rigidity is gradual. The first regions to become flexible in this simulated thermal denaturation are the CP ends of H4, H5, and H6. From state A through the transition state identified as B (Figure 3.2) to the stability core indicated by C, there is a more dramatic loss of rigidity. These three highlighted points in the simulated unfolding pathway (A, B, C) are mapped onto the 3D structure of rhodopsin and shown in Figure 3.2 to demonstrate the gradual dissolution of rigidity from the CP region. Figure 3.2 also indicates where these points lie along the curves corresponding to the order parameter (panel a) and specific heat-like quantity (panel b) (Rader et al., 2002). Both quantities are plotted versus the mean coordination number, <r>, indicating the catastrophic change in rigidity as the protein is “unfolded”.

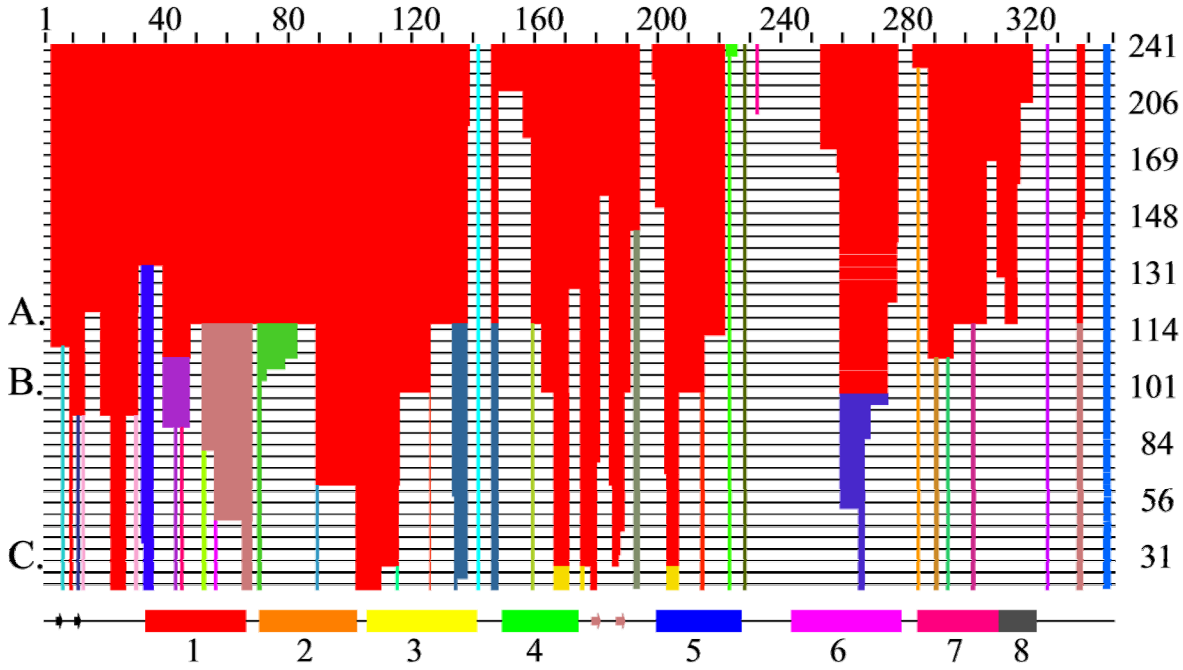

Figure 3.1 Simulated thermal denaturation of rhodopsin
Each line in this hydrogen bond dilution plot depicts which residues are rigid and flexible with a certain set of hydrogen bonds present. Thin black lines represent residues with a flexible backbone while colored blocks identify which rigid cluster residues belong to. As one moves down the hydrogen bond dilution plot, hydrogen bonds are removed one at a time based upon energy. Lines are only shown when there is a change in the backbone rigid clusters. The columns on the left indicate the number of hydrogen bonds (H-bonds) remaining and the mean coordination numbers (<r>). The PDB-defined helical regions are shown as colored blocks and the two β–strands are indicated by arrows along the line immediately below the dilution plot. As described previously (Hespenheide et al., 2002;Rader et al., 2002) removing hydrogen bonds according to energy is analogous to thermal denaturation and hence simulates protein unfolding. Lines labeled A, B and C are representative of a native-like structure, transition state, and folding core, respectively. See Figure 4 and the text for the definitions of the transition state and folding core (Rader et al., 2004).

The core (C) consists of parts of H3 H4 and H5; the β–sheet in the EC2; and some of EC2 and EC3. It includes atoms in residues 9, 10, 22-27, 102-116, 166-171, 175-180, 185-188, 203-207, and 211. Both Cys110 and Cys187 are part of this core and remain mutually rigid even in the final line of the dilution plot. The potential unfolding pathway of rhodopsin presented in Figure 3.1 suggests that the EC ends of H3, H4, and H5 form the most stable region in the

protein. Much of the region identified as the stability core in Figure 3.1C overlaps with the retinal binding region (e.g. Glu113, Ser186, Met207 and His211 are within 4.5 Å of 11-*cis*-retinal). Repeating the simulated denaturation with all protein-retinal interactions removed produced no significant change in the simulated unfolding pathway (data not shown). This suggests the stability core may also serve as a rigid docking site for retinal binding during the folding process.

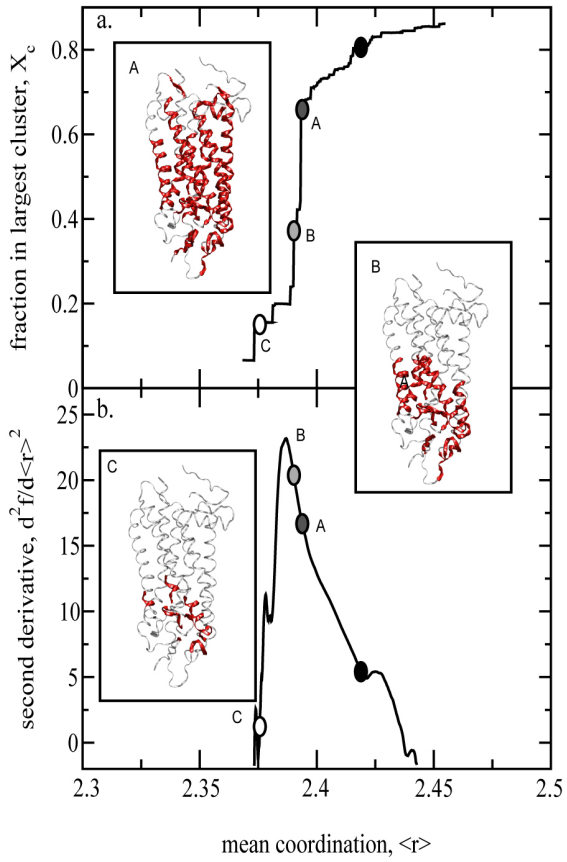

Figure 3.2 Simulated unfolding of rhodopsin. a. Order parameter and b. Specific heat-like curve. The fraction of the number of atoms participating in the largest cluster, Xc, as a function of the mean coordination number, <r>, serves as the order parameter for this system. The peak in the second derivative of the number of floppy modes with respect to <r> is used to identify the transition state (B). Snapshots of the 3D structure, drawn with VMD (Humphrey et al., 1996), are shown along the unfolding pathways, corresponding to highlighted positions A,B, and C in Figure 3.1. The largest rigid cluster, shown by red ribbons, is seen to decrease as the protein unfolds from the native-like state (A) through the transition state (B) to the folding core (C) (Rader et al., 2004).

3.4.2 Peak residues in high frequency modes

The high frequency end of the spectrum, is usually viewed as "uninteresting" and eliminated in NMA or the principal component analysis of MD trajectories. The results depend on the range of interactions considered in the analysis because the high frequency modes are highly localized. Previous studies have shown that the peaks in the fastest GNM modes exhibit some correlation with experimentally determined folding nuclei. Figure 3.3 plots the mode shapes averaged over the fastest ten modes for two cutoff distances: r_c = 10Å (black, solid) and r_c = 7Å (gray, dotted). The results depend on the range of interactions considered in the analysis because the high frequency modes are highly localized. Increasing r_c from the typical distance of 7Å to 10Å better captured the cooperative process of protein folding by identifying residues participating in slightly larger local clusters and gave a set of residues more commensurate with the FIRST folding core residues. Using r_c = 10Å, sixteen peak residues are identified (residues Asn83, Met86-Val87, ***Phe103***, ***Cys110***, ***Glu113-Gly114***, A117, ***Arg177***, ***Pro180***, Met183, Gln184, ***Cys185***, ***Ser186***, ***Cys187***, and Ala299). Included in this set are the cysteine residues Cys110 and Cys187 that form the critical disulfide bond, nine residues identified by both FIRST and GNM (listed in bold and italic and shown by colored tubes in Figure 3.4), and other residues at H2 and H3. The correlation between a majority of these residues and the core residues identified by FIRST suggests that these residues are structurally important for stabilizing the folded state.

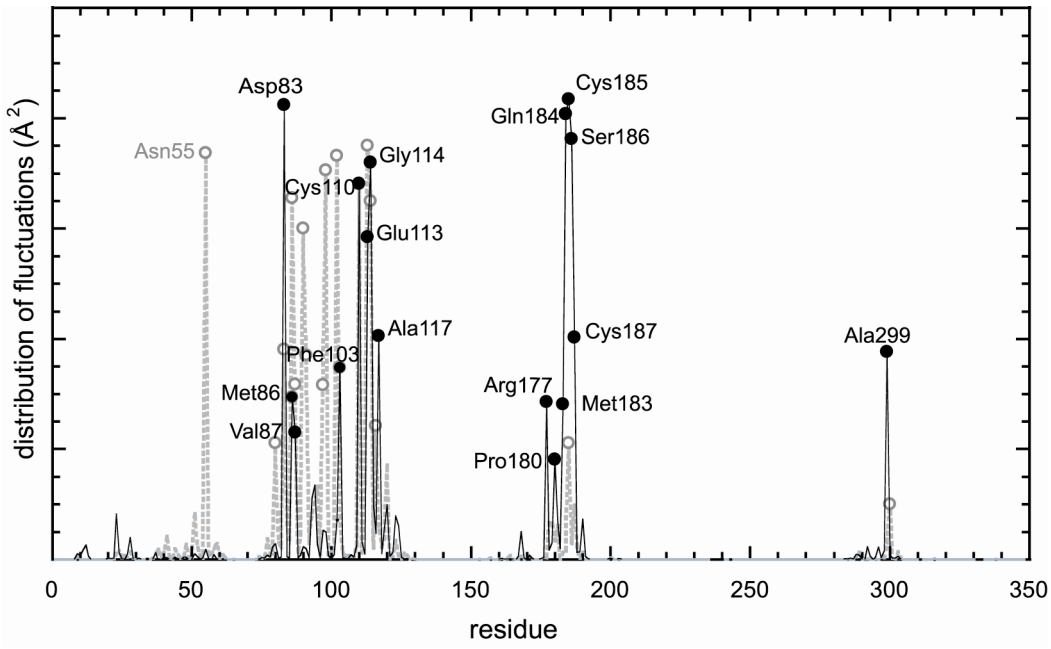

Figure 3.3 Distribution of fluctuations in the high frequency modes of GNM
The solid curve with filled circles displays the mode shape averaged over the fastest ten GNM modes calculated with r_c = 10 Å. The peaks indicate the residues that are most likely to participate in the folding nucleus of rhodopsin. 16 residues with values above a threshold of $6N^{-1}$, where N = 348 residues, are labeled along with the dominant peak from r_c = 7 Å at Asn55. The dashed gray line refers to the results with r_c = 7Å which indicate more localized fluctuations. Corresponding additional peaks (not labeled) are at Ala80, Asn83, Phe91, Thr97, Ser98, and Tyr102.

3.4.3 Experimental validation of folding core

To test the relevance of these computational results, the residues predicted to be critical for folding by FIRST and the peaks in the highest frequency modes of GNM were compared to mutational data on the folding of rhodopsin. A dataset was extracted from the literature containing point mutations and deletions of four or fewer residues. Mutations and deletions were considered to cause no (i.e. folding), partial, or complete misfolding depending on the mutant's ability to bind retinal(Karnik et al., 1988). Table 1 summarizes the comparison of the predicted folding core with experimental data (Karnik et al., 1988;Anukanth and Khorana, 1994;Kaushal et al., 1994;Liu et al., 1996;Hwa et al., 1999;Ridge and Abdulaev, 2000). Mutagenesis experiments

have been performed for 164 residues, 91 of which did not affect folding, while 73 caused misfolding (40 complete, 33 partial). Present computations indicate a total of 52 residues to play a key role in underlying the stability/folding of rhodopsin, 39 of which can be tested against experiments. 35 out of these 39 residues are confirmed by experiments to cause misfolding, 26 of which complete, and 9 partial. The computational results are presented as three sets: FIRST finds 45 stability core residues colored red in Figure 3.2C, GNM yields 16 fast mode peaks shown in Figure 3.3, and the combination of the two sets of results. The union of these two data sets identifies 52 key residues. Of the residues having experimental data, both computational methods correctly predict over 78% to cause misfolding (complete and partial). 31 of the 34 (91%) FIRST stability core residues are known to cause misfolding. This degree of correlation is remarkable. Additionally, 26 of these 31 residues (83.4%) are known to cause complete misfolding suggesting that the residues identified by FIRST are very significant for initiation of proper folding.

Table 3.1 Comparison of experimental data and computational predictions

Experiment	Number of residues experimentally studied	164		
	Fraction observed not to affect folding	91/164		
	Fraction observed to cause misfolding	73/164		
	Fraction that cause complete misfolding	40/73		
Computations		FIRST	GNM	Both
	Number of key residues theoretically identified	45	16	52
	Fraction also studied by experiments	34/45	14/16	39/52
	Fraction correctly observed to cause misfolding	31/34	11/14	35/39
	Fraction of them that cause *complete* misfolding	26/31	5/11	26/35

3.5 DISCUSSION

3.5.1 Residues important for rhodopsin stability

Figure 3.4A identifies the most stable residues found by FIRST (red) and GNM (green) analysis. The disulfide bond forming cysteine residues are shown in yellow and the other seven residues common between the methods in cyan. Although both methods assume that the information about the folding process is encoded in the native conformation, the sets of stable residues obtained by the two methods differ. The GNM adopts an elastic network description, and provides insights about the functional motions near the native state. FIRST simulates unfolding by systematically altering the network to mimic unfolding and re-analyzes the structure at each step. The GNM fast mode shapes are not as robust as the slow mode shapes, and core residues predicted from fast mode peaks are usually interpreted in conjunction with complementary data from experiments or more detailed simulations. Combining the results from GNM and FIRST software identifies 52 stabilizing residues that spread-out from the disulfide bond and 11-*cis*-retinal binding pocket. This observation is strongly supported by previous experimental evidence, and by the degree of conservation of amino acids in the GPCR family. Experiments have demonstrated that the disulfide bond plays a critical role in the folding and stability of rhodopsin (Anukanth and Khorana, 1994;Liu et al., 1996;Kono et al., 1998;Hwa et al., 1999;Hwa et al., 2001) and examination of the rhodopsin crystal structure suggests that it serves as an anchor for many interactions between 11-*cis*-retinal in the TM domain and the tertiary structure in the EC domain (Okada and Palczewski, 2001). The region in rhodopsin containing the Cys110-Cys187 disulfide bond and the 11-*cis*-retinal chromophore binding pocket is the most rigid, most severely constrained on a local scale and most resistant to thermal denaturation. The

local environment for the nine core residues predicted by both GNM and FIRST analysis is shown in Figure 3.4B. They are positioned at the TM-EC interface of the 11-*cis*-retinal binding pocket.

The majority of GPCRs have cysteines in positions equivalent to Cys110 and Cys187 (Rader et al., 2004) and have a disulfide bond between them. Combining the high degree of conservation with the observed stabilizing role in rhodopsin, it is likely that this disulfide bond couples the EC and TM domains for GPCRs in general. The peaks Asn55, Asn83 and Cys110 in Figure 3.3 are consistent with the conserved residues at H1-H3 suggesting that the conservation of these residues originates, at least in part, from stability/folding requirements. Functional criteria also play an important role. It is conceivable that the interactions at the global hinge site are finely tuned and conserved to comply with the mechanical constraints imposed by the collective dynamics of the protein. Residues Ala80 and Ala299, also peaks in Figure 3.3, are reported to be 60-80% conserved (Mirzadegan et al., 2003). Other peaks observed in Figure 3.3 include residues not expected to be generally conserved in the GPCR family, namely those that coordinate 11-*cis*-retinal. These includes the Schiff base counter ion, Glu113, and its close neighbors at H3 Thr97-Ser98, Leu112, G114, Phe116-Ala117, Gly120 and EL2 including the fourth EC β–strand (Pro180, Met183-Gln184-Cys185-Ser186-Cys187), and Tyr102-Phe103 in EL1. The tight packing near the 11-*cis*-retinal binding site is presumably instrumental for the efficient transmission of light-induced conformational changes to the CP surface.

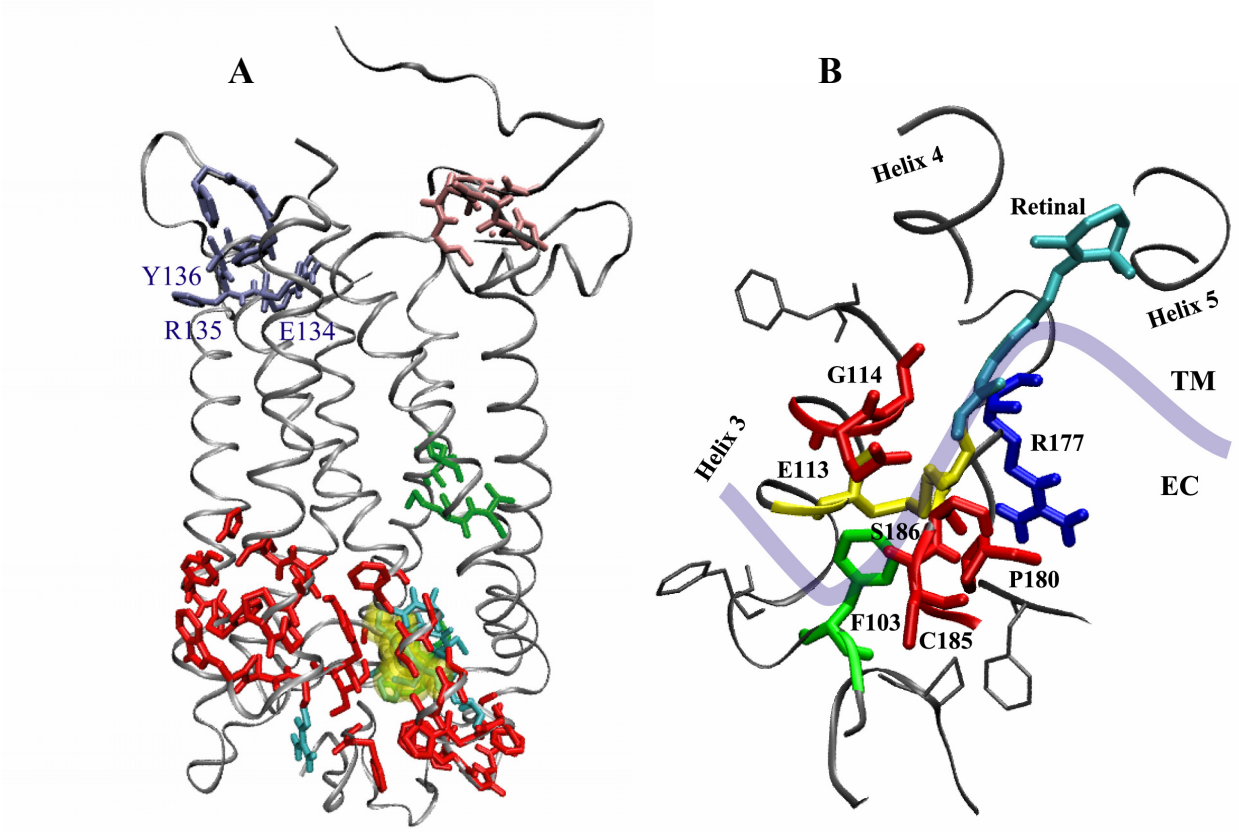

Figure 3.4 Critical folding residues in rhodopsin

(A) FIRST and GNM core residues. Both methods select the cysteine residues forming the critical disulfide bond, shown by space-filling yellow spheres. The other seven residues found in common by these methods are shown in cyan. The remaining seven GNM fast mode peak residues are shown in green and the remaining FIRST core residues from line C in Figure 3.2 are shown in red. The two next largest rigid clusters, from FIRST analysis, are shown in violet and pink to lie at the CP end of the TM region. (B) Local neighborhood of the most stable residues. The nine residues in common between FIRST and GNM fast mode peaks (cyan and yellow in Figure 3.4A) are shown by thick, colored sticks. At the center of this cluster are the cysteine residues (Cys110, Cys187 in yellow) that form the critical disulfide bond. Phe103 (green), five of the common residues (Glu113, Gly114, Pro180, Cys185, Ser186 in red), and the retinal chromophore (retinal, cyan) are located within 4 Å of this disulfide bond and span the TM-EC interface (suggested by the thick blue curve). The other residue in common between these methods, Arg177 (blue), demonstrates how this local stability is propagated across side chain interactions. Side chains for a few of the 45 FIRST folding core residues are shown by thin sticks to orient the reader.

The CP ends of the helices are found by the FIRST analysis to be composed of small, independently rigid groups some of which are present in the transition state and form tertiary contacts that persist even at later stages of unfolding. Two examples of such tertiary contacts shown in pink and blue in Figure 3.4A contain residues known to be highly conserved in GPCR class A, including the D(E)RY motif residues Glu134-Arg135-Tyr136. These rigid clusters are separated by flexible regions from the largest rigid cluster suggesting that the CP domain retains some flexibility despite being structured into small rigid elements. This type of flexibility may be functionally required to undergo the conformational changes that are recognized by other molecules in the signaling cascade, such as the G protein.

4.0 THE MECHANISM OF SIGNAL PROPAGATION UPON ISOMERIZATION: INSIGHTS FROM MOLECULAR DYNAMICS SIMULATIONS GUIDED BY NORMAL MODES

4.1 ABSTRACT

Here, we aim at exploring the functional changes in conformation and the signal transmission mechanisms involved in the photoactivation of rhodopsin in atomic detail in the presence of explicit membrane and water environment. To this aim, we propose a MD simulation protocol that utilizes the normal modes derived from ANM to define the target conformations. The deformations along multiple low frequency modes of motions are used therein to efficiently sample collective changes in conformation. The resulting conformations are compared with experimental data on activation including site-directed mutagenesis, cysteine cross-linking, and site-directed spin labeling analyses. We identified two highly stable regions, one clustered near the chromophore, and the other near the CP ends of TM helices H1, H2 and H7. The CP ends of the helices (H3-H6) critical for G-protein binding and activation are shown to undergo distinctively high mobilities due to the redistribution of interactions in the neighborhood of the chromophore upon stabilization of the *trans* form. Due to tight packing near the chromophore, the local structural rearrangements in the adjoining H3-H6 residues are efficiently propagated to the CP end of these particular helices. At the proposed reconfigured form of rhodopsin, the all-*trans*-retinal interacts with Cys167 on H4 and Phe203 at H5, which were not accessible in the

dark state, and exhibits relatively stronger interactions with H5 residues, while some of the contacts made (in the *cis*-form) with H6 residues are lost in the new conformation.

4.2 INTRODUCTION

In the previous chapters, we adopted a purely analytical approach and examined the collective motions predicted by GNM (Bahar et al., 1997) and ANM (Doruker et al., 2000; Atilgan et al., 2001), followed by energy minimization to optimize side chain conformations. We combined this analysis with a systematic Cys->Ser replacement experiments to develop a model for providing insights into the potential reconfiguration of rhodopsin accompanying its activation (Isin et al., 2006). We further found the residues that are important for stability by using FIRST in conjuction with the high frequency modes of GNM (Rader et al., 2004).

Here, our aim is to find at atomic detail the biologically relevant conformations of rhodopsin, which couple retinal isomerization to conformational changes in both the TM domain and the critical G-protein binding sites on the CP surface. We would like to explore the global dynamics, while incorporating the effects from residue specificities and interactions with lipid and water molecule at atomic details. To this aim we propose an algorithm, referred to as **ANM-restrained MD**, which uses the deformations derived from ENM analysis as restraints in MD trajectories. This permits us to sample the collective motions that are otherwise beyond the range of conventional MD simulations. By this new approach, we aim at incorporating the realism and accuracy of MD into ENM analysis while taking advantage of ENM results to accelerate MD simulations. Remarkably, with this method, we identify a hinge site, which does not change with

several rounds of applying normal modes as restraints. This hinge site includes residues that are directly affected by the isomerization of retinal, as well as those stabilizing the resulting all-*trans* conformation of the retinal. The CP ends of the helices H3, H4, H5, and H6 and connecting loops are found to enjoy an enhanced mobility facilitated by this hinge site. Several new interactions are observed, which permit us to gain a better understanding of the mechanism of signal propagation from the chromophore binding pocket to the G-protein binding sites on the CP domain.

4.3 METHOD

4.3.1 Molecular dynamics

4.3.1.1 Equation of motion

Molecules are quantum mechanical entities and their temporal behaviors are best described by the quantum mechanical equation of motion (the time-dependent Schrödinger Equation). However, this equation is extremely difficult to solve and not feasible for large systems. Most computational studies of biomolecules use atomistic models, in which atoms are the smallest particles in the system rather than the electrons and nuclei used in quantum mechanics. Today's MD simulations rely on the Born-Oppenheimer approximation. Since the nuclei are much heavier than electrons, this approximation assumes that the electronic motion follows the nuclear one and electrons can adjust nuclear motions instantaneously. Conventional MD simulations mostly use the simple classical mechanics and approximate the motion executed by atoms to Newton's equation of motion.

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \ddot{\mathbf{r}}_i \tag{4.1}$$

where $\mathbf{F}_i$ is the force acting on particle i, m_i is the mass of particle i, $\mathbf{a}_i$ is its acceleration, and $\ddot{\mathbf{r}}$ is the second time derivative of the particle position $\mathbf{r}$. The force $\mathbf{F}_i$ is determined by the gradient of the potential energy function, $V(\mathbf{r})$, which is a function of all the atomic coordinates $\mathbf{r}$ (Leach, 2001;Schlick, 2002).

4.3.1.2 Potential energy functions

A potential energy function is a mathematical equation that allows for the potential energy, V, of a chemical system to be calculated as a function of its three-dimensional (3D) structure. The equation includes terms describing the structural and chemical properties system. The energy of a biomolecular system is additive, hence it can be expressed as a sum of potentials derived from physical forces: van der Waals, electrostatic, strains arising from ideal bond length and angle deviations, and internal torsion flexibility (MacKerell, Jr., 2004;van Gunsteren et al., 2006). The forces can be separated into bonded (internal) and nonbonded (external) terms.

$$V_{total} = V_{bonded} + V_{nonbonded} \tag{4.2}$$

4.3.1.3 Bonded interactions

The potential function for the bonded interactions contains terms for bonds, angles and dihedrals:

$$V_{bonded} = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi \left[1 + \cos(n\chi - \sigma)\right] \tag{4.3}$$

where b are the instantaneous bond lengths and θ are the instantaneous angles between two bonds sharing an atom in the middle. χ denotes the dihedral angle and n is an integer which refers to the periodicity of the rotational barrier (MacKerell, Jr., 2004). A schematic representation of two molecules is shown in Figure 4.1 to illustrate the terms calculated in potential energy function. Bonded potentials can be considered as strain terms that model small scale deviations about reference values, b_0, θ_0, and, σ.

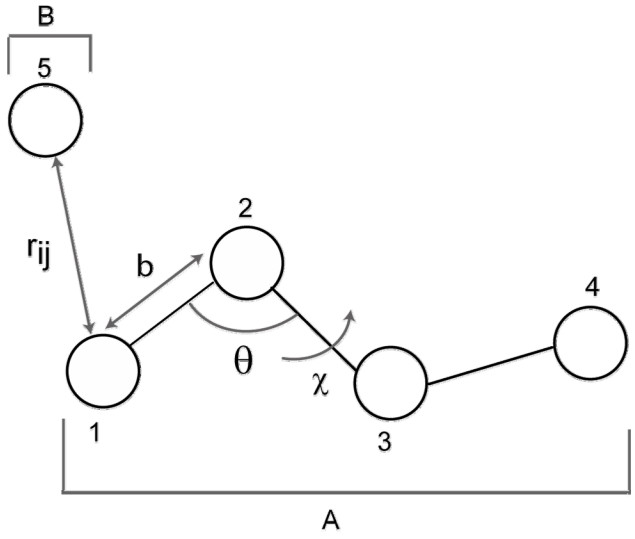

Figure 4.1 Schematic representation of two interacting molecules
Molecule A consists of 4 atoms (1, 2, 3, and 4) and molecule B has one atom (5). Potential energy terms taken from the 3D structure include bond lengths (b), bond angles (θ), dihedral angles (χ) and distances between atoms (r_{ij}).

4.3.1.4 Nonbonded interactions

Nonbonded interactions are considered to be the most important contributors to the potential energy function of the biological systems. They are critical for accomplishing long-range interactions within the molecules as well as stabilizing biomolecular conformations in the solvent

environment. Nonbonded interactions are represented by two different potentials: van der Waals and Coulombic (Adcock and McCammon, 2006).

4.3.1.5 The van der Waals potential

The best known van der Waals potential function is the Lennard-Jones potential. It consists of attractive and repulsive parts:

$$V_{Lennard-Jones} = \sum_{\substack{nonbonded \\ atom\ pairs}} \varepsilon_{ij} \left[(\frac{A_{ij}}{r_{ij}})^{12} - (\frac{B_{ij}}{r_{ij}})^{6} \right] \tag{4.4}$$

where the attractive (A) and repulsive coefficients (B) depend on the type of the two interacting atoms. The strong tendency of atoms to repel one another when their electron clouds overlap is represented by r^{-12}. As the distance between two atoms approaches to infinity the Lennard-Jones potential decays rapidly to zero. Hence, the van der Waals energies and forces are short range and they are usually computed for pairs of atoms only within a certain cutoff distance. The attraction between two atoms as they approach an optimal internuclear distance is represented by r^{-6}.

4.3.1.6 The Coulomb potential

The ionic interactions between fully or partially charged pairs of atoms can be represented by by the Coulomb's law

$$V_{coulomb} = \sum_{\substack{nonbonded \\ atom\ pairs}} \frac{q_i q_j}{\varepsilon_D r_{ij}} \tag{4.5}$$

where $V_{coulomb}$ is the Coulomb potential, q_i and q_j are the respective charges of atoms i and j, r_{ij} is the distance between i^{th} and j^{th} atoms. ε_D is dielectric constant. The Coulomb potential decay's with distance much slower than the Lennard-Jones and it plays a bigger role in stabilizing the long range interactions (Schlick, 2002).

4.3.1.7 Parameterization and force fields

The potential energy function is composed two different groups of terms. The first group of terms can be directly obtained from the 3D structure of the molecule as illustrated in Figure 4.1. The second group contains empirical parameters which are associated with the particular types of the atoms. A force field defines all parameters for each type of atom as well as potential energy functions governing the interactions between all atom types that are present in the biological system. The optimization of force field parameters is necessary to ensure that calculations reproduce experimental findings for vibrational frequencies, heats of formation, intermolecular energies, geometries, torsional barriers, etc. The quality of the parameters is crucial for the accuracy of the results obtained from computational studies of biological molecules. Currently, a variety of force fields are available for the simulations of biomolecules. Major force fields are CHARMM (MacKerell et al., 1998), AMBER (Pearlman et al., 1995;Case et al., 2005) and GROMOS (Scott et al., 1999). While their respective potential energy equations are very similar, these force fields do differ in the optimization procedures of the parameters (MacKerell, Jr., 2004;van Gunsteren et al., 2006;Adcock and McCammon, 2006). Most of these simulations are performed by software packages such as NAMD (Phillips et al., 2005), GROMACS (Vandrunen et al., 1995;Van der Spoel et al., 2005), CHARMM (MacKerell et al., 1998) and AMBER

(Pearlman et al., 1995;Case et al., 2005). These softwares can work with potential functions, parameters and file formats of different force fields.

4.3.1.8 Force

The force $\mathbf{F}_i$ is determined by the negative gradient of the potential energy function, $V(\mathbf{r})$, which is a function of all the atomic coordinates $\mathbf{r}$,

$$\mathbf{F}_i = -\nabla_i V(\mathbf{r}) \tag{4.6}$$

4.3.1.9 Periodic boundary conditions

In a simulation system, biomolecules and surrounding solvent molecules are placed in a unit cell and considered to have infinitely many images in space. This replication forms an infinite lattice in 3D space. Regardless of the position within the central cell, an atom is surrounded on all sides by other atoms or by their images. When a particle leaves the cell form one side, it is replaced by an identical copy which enters from the opposite side. In such periodic boundary conditions, each particle in the system is subject to a potential exerted by all other particles including those in the surrounding image cells. This eliminates entirely the surface effects. Unlike finite boundary conditions, there is no vacuum interface and no specific boundary potentials are necessary (Leach, 2001;Allen and Tildesley, 2006).

4.3.1.10 The Ewald's sum

Truncation of nonbonded interactions results in severe artifacts due to the large increase in the forces governing these interactions the truncation boundary. The Ewald's method has been developed to eliminate the discontinuity arising from truncated long-range forces. It calculates the electrostatic energy of a system with periodic conditions (Essmann et al., 1995;Cheatham et

al., 1995). The total electrostatic energy of a system with periodic boundary conditions is written as:

$$V_{electrostatic} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\mathbf{n}_r}^{'}\frac{q_i q_j}{\left|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r\right|} \tag{4.7}$$

where q_i is the charge on atom i and $\mathbf{r}_i$ is its position vector. N is the number of charges and $\mathbf{n}_r$ is the lattice vector and it is represented as $\mathbf{n}_r = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ where $\mathbf{a}_1, \mathbf{a}_2$ and $\mathbf{a}_3$ are base vectors and n_1, n_2 and n_3 are constants. The prime indicates that terms with $i = j$ and $\mathbf{n}_r = 0$ are omitted. This prevents the self interaction but atoms still interact with their replica images. The total electrostatic energy function (equation 4.7) is only conditionally convergent. The Ewald's sum is the reorganized form of this function into sums over concentric spherical shells and it converges rapidly by using the following erfc function:

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,dt \tag{4.8}$$

The Ewald's sum includes a standard direct sum for short range interactions, a reciprocal sum for long range interactions, a self energy and a surface energy terms:

$$E_{Ewald} = E_{direct} + E_{reciprocal} + E_{self} + E_{surface} \tag{4.9}$$

$$E_{direct} = \frac{1}{2}\sum_{i,j=1}^{N} q_i q_j \sum_{\mathbf{n}_r}^{'}\frac{\mathrm{erfc}(\beta\left|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r\right|)}{\left|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r\right|} - \sum_{(i,j)\in\text{excluded}}\frac{q_i q_j}{\left|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r\right|} \tag{4.10}$$

$$E_{reciprocal} = \frac{1}{2\pi V} \sum_{\vec{m} \neq 0} \frac{\exp(-\pi^2 |\mathbf{m}|^2 / \beta^2)}{|\mathbf{m}|^2} \left| \sum_{i=1}^{N} q_i \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_i) \right|^2 \quad (4.11)$$

$$E_{self} = -\frac{\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \quad (4.12)$$

$$E_{surface} = \frac{2\pi}{(2\varepsilon_s + 1)} \left| \sum_{i=1}^{N} q_i \mathbf{r}_i \right|^2 \quad (4.13)$$

The rearranged sum includes an adjustable parameter β, whose value must be determined to maximize numerical accuracy. It shifts the computational load between direct and reciprocal sums. V is the volume of the simulation box. ε_s is the dielectric constant of the solvent. $\mathbf{m}$ is the reciprocal vector and it equals to $\mathbf{m} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3$. m_1, m_2, and m_3 are integers. $\mathbf{b}_1$, $\mathbf{b}_2$, and, $\mathbf{b}_3$ are base vectors. Multiplication of lattice and reciprocal vectors gives:

$$\mathbf{a}_\alpha \cdot \mathbf{b}_\alpha = \delta_{\alpha\beta} \quad \text{where } \alpha, \beta = 1, 2, 3. \quad (4.14)$$

4.3.1.11 Computational algorithms for MD trajectories: Taylor expansion and Verlet integrators

To generate MD trajectories of a biomolecular system containing several particles, Newton's equation of motions are integrated numerically by finite difference techniques. The integration is discretized into small time steps that are significantly smaller than the motion of interest. The continuous variables $\mathbf{r}(t)$ and $\mathbf{v}(t)$ are approximated at each discrete time intervals. In integration algorithms, the positions and dynamic properties such as velocities and accelerations are approximated by Taylor series expansions. The most common algorithm used in biomolecular studies is the Verlet algorithm (Allen and Tildesley, 2006). To obtain new

positions at $t+\Delta t$, the Verlet algorithm uses positions and accelerations at times t and $t-\Delta t$ in a Taylor expansion which can be written for $\mathbf{r}(t+\Delta t)$ at time $t+\Delta t$ as:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\Delta t^2 + ... \tag{4.15}$$

or

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + ... \tag{4.16}$$

where $\mathbf{v}(t)$ is the velocity vector and $\mathbf{a}(t)$ is the acceleration. The equation 4.15 written can be in a discrete form as

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n\Delta t + \mathbf{a}_n\Delta t^2 + O(\Delta t^3) \tag{4.17}$$

where $\mathbf{r}_{n+1}$ and $\mathbf{r}_n$ are the positions at steps $n+1$ (at time $t+\Delta t$) and n (at time t) respectively. $\mathbf{v}_n$ is the velocity and $\mathbf{a}_n$ is the acceleration at time t.

Similarly, $\mathbf{r}_{n-1}$ at $t-\Delta t$ is

$$\mathbf{r}_{n-1} = \mathbf{r}_n - \mathbf{v}_n\Delta t + \frac{1}{2}\mathbf{a}_n\Delta t^2 + O(\Delta t^3) \tag{4.18}$$

Equations 4.17 and 4.18 are added to obtain $\mathbf{r}_{n+1}$ as functions of $\mathbf{r}_n$, $\mathbf{r}_{n-1}$ and $\mathbf{a}_n$:

$$\mathbf{r}_{n+1} = 2\mathbf{r}_n - \mathbf{r}_{n-1} + \frac{1}{2}\mathbf{a}_n\Delta t^2 + O(\Delta t^4) \tag{4.19}$$

By using equation 4.19, the following algorithm is performed,

1. Calculate $\mathbf{F}_n$ and $\mathbf{a}_n$ by using $\mathbf{r}_n$ and equation 4.1
2. Find $\mathbf{r}_{n+1}$ by inserting $\mathbf{r}_{n-1}$, $\mathbf{a}_n$ (found at the step 1) and $\mathbf{r}_n$ to equation 4.19.

Although the main equation 4.19, does not contain the velocity term, the velocity can be obtain by subtracting the equation 4.18 from 4.17:

$$\mathbf{v}_n = \frac{\mathbf{r}_{n+1} - \mathbf{r}_{n-1}}{2\Delta t} + O(\Delta t^2) \tag{4.20}$$

However, the lack of an explicit velocity term in the main equation solved by the Verlet algorithm is a disadvantage. The velocities at step n (time t) can only be calculated by equation 4.20, after the new position $\mathbf{r}_{n+1}$ is calculated by using $\mathbf{r}_n$. Another disadvantage is adding a very small term ($\mathbf{a}(t)\Delta t^2$) to the difference of two larger terms, $2\mathbf{r}(t)$ and $\mathbf{r}(t-\Delta t)$ to obtain $\mathbf{r}(t+\Delta t)$. This leads to the loss of precision.

A variant of the verlet algorithm, the leap-frog algorithm, explicitly includes the velocity term and also does not require to take the difference of two large numbers. It uses the following equations:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_{n+1/2}\Delta t \tag{4.21}$$

$$\mathbf{v}_{n+1/2} = \mathbf{v}_{n-1/2} + \mathbf{a}_n\Delta t \tag{4.22}$$

where $\mathbf{v}_{n+1/2}$ and $\mathbf{v}_{n-1/2}$ are the respective velocities at time steps $t_{n+1/2}$ and $t_{n-1/2}$. The leap-frog algorithm follows these steps:

1. Calculate $\mathbf{F}_n$ and $\mathbf{a}_n$ by using $\mathbf{r}_n$.
2. Calculate next mid-step velocity $\mathbf{v}_{n+1/2}$ by inserting $\mathbf{F}_n$ and $\mathbf{v}_{n-1/2}$ into equation 4.22.
3. Calculate next step position $\mathbf{r}_{n+1}$ by inserting $\mathbf{r}_n$ and $\mathbf{v}_{n+1/2}$ into equation 4.21.

Although the leap-frog algorithm overcomes the disadvantages of Verlet algorithm, it has its own disadvantages. Its velocity and positions are not synchronized. Hence, it is not possible to

calculate the kinetic energy contribution to the total energy at the time when the positions are defined.

Another variant of Verlet algorithm, velocity Verlet, calculates positions, velocities, and accelerations at the same time t and also minimizes the round-off errors. Hence, it provides more accurate computation of velocities and kinetic energies of the system than Verlet and leap-frog. The velocity Verlet algorithm uses the following equations and algorithm:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n \Delta t + \frac{1}{2}\mathbf{a}_n \Delta t^2 \tag{4.23}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \frac{1}{2}(\mathbf{a}_n + \mathbf{a}_{n+1})\Delta t \tag{4.24}$$

$$\mathbf{v}_{n+1/2} = \mathbf{v}_n + \frac{1}{2}\mathbf{a}_n \Delta t \tag{4.25}$$

$$\mathbf{v}_n = \mathbf{v}_{n+1/2} + \frac{1}{2}\mathbf{a}_{n+1} \Delta t \tag{4.26}$$

1. Calculate the position $\mathbf{r}_{n+1}$ using equation 4.23.
2. Calculate $\mathbf{v}_{n+1/2}$ using equation 4.25.
3. Calculate $\mathbf{F}_{n+1}$ and $\mathbf{a}_{n+1}$ at time $t+\Delta t$.
4. Calculate $\mathbf{v}_n$ by inserting $\mathbf{v}_{n+1/2}$ and $\mathbf{a}_{n+1}$ to equation 4.26.

The only disadvantage of the velocity Verlet algorithm comes from its more complex equations which make the algorithm computationally more expensive.

4.3.1.12 Simulation steps in a molecular dynamics study

4.3.1.12.1 Setting up the biomolecular system and assigning the initial values of the system

The Initial coordinates of molecules are usually taken from structures obtained by using X-ray crystallography or NMR experiments, deposited in the PDB. In the absence of an experimental structure, models are generated by computational structure determination techniques (Sanchez et al., 2000). Most experimentally found structures miss certain atoms such as hydrogens or even some highly flexible/disordered segments that might be important for the characteristic motions of the system studied. After modeling the missing parts computationally, the biomolecule is solvated in water or embedded into a membrane. Initial coordinates of solvent and lipid molecules are typically obtained from pure solvent and membrane simulations. Salt ions are added to adjust the charge concentration of the system to the desired value. Initial coordinates are then subjected to energy minimization to relieve possible artificial contacts and steric clashes that may have occurred during the model construction as well as to correct the possible inaccuracies in bond lengths and bond angles that exist in the experimental structure. Energy minimization techniques will be explained in detail in the following chapters (Allen and Tildesley, 2006).

4.3.1.12.2 Initial velocities

The only useful information for setting initial velocities is the simulation temperature. According to the equipartition theorem, the absolute temperature of a particular system is defined by its kinetic energy as

$$T(t) = \frac{1}{k_B N_F} \sum_{i=1}^{N} m_i \mathbf{v}_i^{\,2} \tag{4.27}$$

where N_F is the total numbers of degrees of freedom in the system, N is the number of atoms and k_B is the Boltzmann constant (Schlick, 2002).

The initial velocities are thus assigned in accord with the total kinetic energy of the system at the simulation temperature using a Gaussian (or Maxwellian) velocity distribution. Due to the randomness of velocities (that obey to the given Gaussian distribution), atoms with high velocities may cluster at certain regions of the system. Additionally, no correlation can exist between the initial velocities of the neighboring atoms. These artificial effects arising from random assignment of velocities may lead to an unstable system. To overcome this problem, simulations usually start by assigning velocities at a low temperature which are then gradually increased to conform to the desired simulation temperature (Schlick, 2002).

4.3.1.12.3 Choosing the time step

The time step is limited by both accuracy and stability. Properties of Newton's equation of motion are used to ensure that the numerical solution is correct and the system is stable. These properties are the conservation of energy, linear and angular momentum. The potential energy function does not depend explicitly on time or velocities; it only depends on the particle separation and orientations. Therefore, in the absence of an external field, the energy, and the linear and angular momentum terms must remain constant. Instabilities lead to a violation of energy, and linear and angular momentum conservation principle, and may eventually result in a program failure.

The Newton's equation of motion is reversible in time. This is indeed another property that is used to examine the accuracy of the generated trajectories. In principle, if the equations of motion are solved correctly, the trajectory should be retraced by changing the sign of velocities. However, large molecular systems are generally very chaotic, and the time reversibility of their trajectories is practically valid for only short periods of time.

While a simulation is stable and produces numerically 'correct' data, the time step may be too large to capture high-frequency motions. Hence, accuracy requirements limit the time step. The resolution of the trajectories, or the step size, must be adequate for the nature of the process simulated and for a useful analysis of the results.

4.3.1.13 Equilibration

Once the initial configuration of the system is generated, an equilibration phase is simulated before the collection of data in the production phase. During equilibration, a set of thermodynamic quantities are monitored such as the fluctuations in energy, temperature and pressure as well as those in some structural properties of the system. Equilibration continues until the monitored properties converge to their equilibrium properties (Allen and Tildesley, 2006).

4.3.1.14 Production and data analysis

During the production phase, snapshots of coordinates and velocities are saved for each particle at regular time intervals, which constitute the trajectory of the system. Then, these data are analyzed not only to extract information about the system but also to ensure that the simulations run properly and give expected values for certain quantitites. The quantitites analyzed highly

depend on the characteristics of the simulation. The output from an MD simulation is generally taken about once every hundred steps, each step being about 1 or 2 fs. Depending on the simulated system, specific quantities are sampled such as the force on an applied restraint, the particular interaction energies between selected elements of system. The conformational changes being sampled may be examined using visualization tools such as VMD (Humphrey et al., 1996).

4.3.2 Targeted molecular dynamics

Targeted molecular dynamics (TMD) is used to induce conformational transitions from a given initial conformation (*I*) to a known target conformation (*T*) at ordinary temperature by applying a set of time-dependent, geometrical constraints (Figure 4.2). The transition is enforced, independent of the height of energy barriers, while the dynamics of the molecule is only minimally influenced by the constraints (Schlitter et al., 1994). In the TMD method, the instantaneous distance ρ(*t*) of each configuration from the target configuration can be written as:

$$\rho(t) = \left|\mathbf{r}(t) - \mathbf{r}_T\right| = \left(\sum (\mathbf{r}_i(t) - \mathbf{r}_{Ti})^2\right)^{\frac{1}{2}} \tag{4.28}$$

where $\mathbf{r}(t)$ is the conformation vector of the biomolecule and $\mathbf{r}_i(t)$ is the position vector of the atom *i* at time *t*. Likewise, $\mathbf{r}_T$ is the target conformation vector for the biomolecule and $\mathbf{r}_{Ti}$ is the position vector of the atom i at the target conformation.

TMD uses the following constraint to reach the target structure:

$$\Phi(\mathbf{r}(t)) = \sum (\mathbf{r}_i(t) - \mathbf{r}_{Ti})^2 - \rho^2(t) = 0 \tag{4.29}$$

The additional force due to the implementation of this constraint can be written as:

$$\mathbf{F}_c(t) = k\frac{\partial \Phi}{\partial x} = 2k(\mathbf{r}(t) - \mathbf{r}_T) \tag{4.30}$$

where k is the force constant .

A typical TMD run consists of the following steps:

1. Set the distance ρ_I between target and initial conformation ($\mathbf{r}_I$) as $\rho_I = |\mathbf{r}_I - \mathbf{r}_T|$

2. Solve the equation of motion containing the additional $\mathbf{F}_c$ for the constraint, by assigning initial coordinates using the initial conformation and an appropriate set of initial velocities.

3. At each succeeding time step, decrease the distance by $\Delta\rho = (\rho_I - \rho_T)\frac{\Delta t}{t_s}$ where t_s is the simulation time and ρ_T, is the distance of the conformation from the target structure at the end of the simulation, which should be as small as possible. An alternative method for calculating the distance of the conformation at time t from the target is $\rho(t) = \rho(t - \Delta t) - \Delta\rho$.

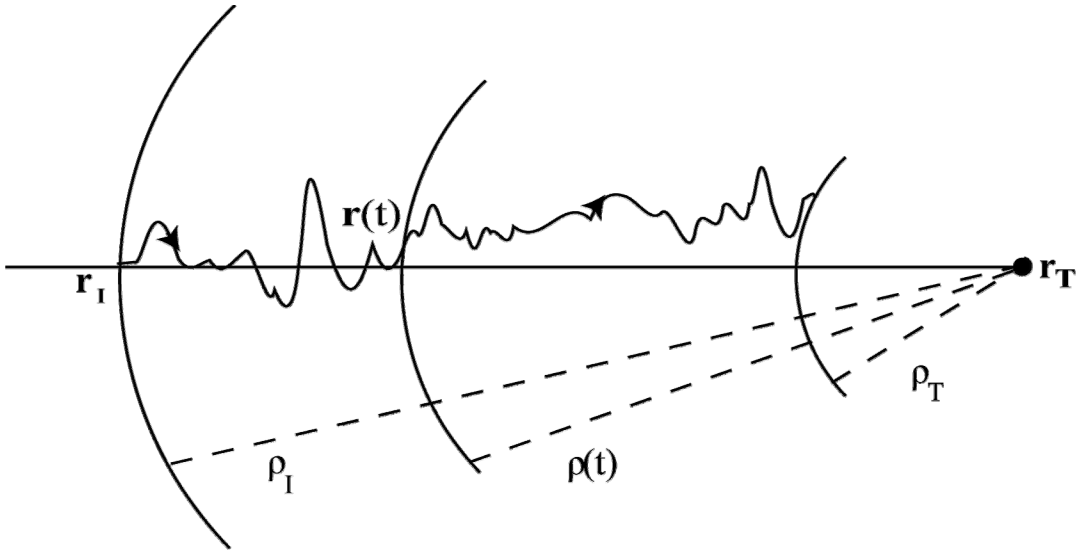

Figure 4.2 Two dimensional representation of a pathway that the molecule follows in a TMD simulation
The distance $\rho(t)$ from the target structure is gradually decreased during the course of simulation. Dashed lines show the distances of the molecule from the target structure at the initial conformation (ρ_I), at time t ($\rho(t)$), and at the final (ρ_T) conformation.

4.3.3 Restrained targeted molecular dynamics

Here, the system is subjected to the following harmonic restraint:

$$T = \frac{k}{2}(\rho(t) - \rho_T(t))^2 \tag{4.31}$$

where $\rho_T(t)$ is the desired distance from the target conformation at time t and $\rho(t)$ is the instantaneous distance from the target. The simulation is divided into sections with different values of $\rho_T(t)$. An alternative approach is to increase the force constant during the simulation.

4.3.4 Energy minimization

Energy minimization methods reduce the potential energy of a given system by relieving local strains in the structure. They are used for many purposes in molecular simulations and modeling. To ensure that the initial structures used in MD and NMA do not contain any high energy interactions, an energy minimization is performed prior to initiation of simulations. Conformations generated during MD simulations often correspond to transient conformations. These conformations are generally subjected to energy minimization to reach a stable conformation at the nearest local minima, before starting to analyze the conformation.

Minimization techniques permit the molecule to go downhill on the multidimensional conformational energy surface. They are, however, unable to surmount energy barriers and they can only locate the local minimum closest to the point from which the minimization process started (unless an algorithm such as Metropolis Monte Carlo is implemented that allows for occasional ascents in the energy surface). Hence, they are not capable of locating the global

energy minimum starting from an arbitrary position. Minima are located by changing the coordinates gradually and producing configurations in the direction which lower the energy of the biomolecule. This procedure is performed until the minimum which is nearest to the starting point is reached. Figure 4.3 shows a schematic energy surface and the minima that would be reached upon initiation of the minimization scheme at different points on the schematic potential energy surface.

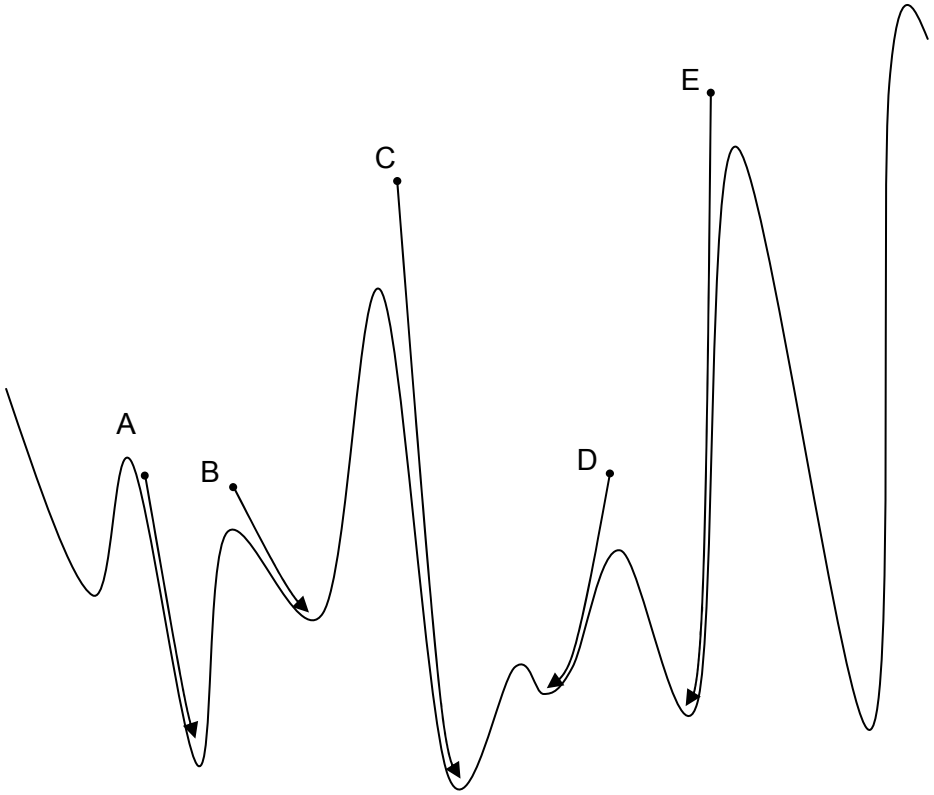

Figure 4.3 Schematic representation of energy minimization
Energy minimization of a structure leads it to the nearest local minima depending on the location of initial structure on the energy surface. The points A, B, C, D, and E represent the starting structure for different energy minimization schemes, which end up in different minima on the energy surface.

The first and second gradients of the potential energy provides information on the slope and curvature of the energy surface, respectively. The first derivative points to the direction where the nearest local minimum lies and its magnitude indicates the steepness of the local slope.

The second derivative gives the curvature of the energy function and are used to predict where the function will change direction (Schlick, 2002).

Minimization methods are classified according to the amount of information that they derive from the potential energy function. Methods that use only the value of the function itself are called zero order minimization methods. Techniques that use derivatives are named as first order and second order methods. First order methods use only the first derivative while second order methods use both first and the second derivatives. The first order methods have a good balance between accuracy and efficiency. Hence, two first order methods, steepest descent and conjugate gradient, have been most commonly used for energy minimization. Both methods use the following equation for the iteration until it converges or it reaches a preset maximum number of steps:

$$\mathbf{r}_k = \mathbf{r}_{k-1} + \lambda_k \mathbf{S}_k \tag{4.32}$$

Here $\mathbf{r}_k$ and $\mathbf{r}_{k-1}$ are the position vectors at steps k and $k-1$ respectively. λ_k is the step size and $\mathbf{S}_k$ is direction at step k (Leach, 2001;Schlick, 2002). Next, these two most common methods are described.

4.3.4.1 Steepest descent.

The direction at each minimization step is directly calculated by the gradient of the energy function:

$$\mathbf{S}_k = -\mathbf{g}_k = -\nabla V(\mathbf{r}) \tag{4.33}$$

where $\mathbf{g}_k$ is the gradient at step k and $U(\mathbf{r})$ is the potential energy function. At each iteration step, the step size is adjusted to compensate the lack of curvature information. When the energy

of the conformation at step k is lower than the one at the preceeding step (k-1), the step size is increased in the following step (k+1).If it is higher than the previous one, the step size at k+1 is decreased to correct the direction of the minimization. Furthermore, SD's imprecise approach to the minimum usually means that the method does not converge to one point. SD gets close to the minimum and stuck around the minimum but rarely reaches it.

4.3.4.2 Conjugated gradients

To calculate the direction at step k, conjugate gradient uses the gradient at k and the direction and gradient of the previous step. The directions of the first step and the successive steps are calculated as follows:

$$\mathbf{S}_1 = -\mathbf{g}_1, \qquad \mathbf{S}_k = -\mathbf{g}_k + b_k \mathbf{S}_{k-1} \tag{4.34}$$

where $b_k = \left| \frac{\mathbf{g}_k}{\mathbf{g}_{k-1}} \right|^2$.

4.3.5 ANM-restrained MD simulations

The idea is to use an ensemble of ANM modes in an iterative scheme, as described in Figure 4.4. Essentially, our proposed algorithm consists of two loops: The first (inner) loop generates a succession of conformations using ANM modes as harmonic restraints in MD runs, succeeded in each case by a short energy minimization algorithm to allow the molecule to settle in a local energy minimum. To this aim, we select from a pool of ANM modes, the subset of the most distinctive (based on eigenvalue dispersion) and cooperative (based on collectivity $\kappa(k)$; see equation 4.35 below) modes. For each mode we define two target conformations (4.36) and we run short MD simulations (of < 0.1 ns) in the presence of harmonic restraints that favor these

target structures. Because the restraints may lead to unrealistic strains in the structure, we perform a short energy minimization succeeding each run and choose among the two alternative structures the one that is energetically favored. After a sufficient number of iterations, by screening all selected modes, a new cycle (outer loop) is initiated with the updated ANM modes corresponding to the structure reached by the end of the first cycle, and this procedure is terminated after a certain number of cycles. The underlying assumption in this protocol is that ANM-derived restraints drive the excursion of the molecule toward a direction that would otherwise be naturally selected at much longer times. Below are more details on the individual steps:

4.3.5.1 Generating the normal modes

We use the ANM for generating the modes of interest. The ANM has been described in detail in the previous chapter.

4.3.5.2 Selection of distinctive and cooperative modes

In principle we select the lowest frequency modes. However, two additional criteria were considered: mode frequency dispersion (or eigenvalue distribution) and degree of collectivity. The former is examined to identify a subset that has distinctive frequencies. The degree of collectivity for the k^{th} mode, on the other hand, is calculated using

$$\kappa(k) = \frac{1}{N}\exp(-\sum_{i}^{N}[\Delta\bar{R}_i(k)]^2 \log[\Delta\bar{R}_i(k)]^2) \tag{4.35}$$

where $\Delta\bar{R}_i(k)$ is the displacement of the i^{th} residue driven by mode k. $\Delta\bar{R}_i(k)$ is normalized such that $\sum_{i}^{N}[\Delta\bar{R}_i(k)]^2 = 1$ (Tama and Sanejouand, 2001) to ascertain that the selected modes are cooperative enough. This criterion helps in eliminating the cases where the low frequency modes

induce a motion in a loosely coupled chain segment only (e.g., the N- or C-terminus). In the case of rhodopsin, the lowest frequency modes were also observed to be the most cooperative ones, and the frequency distribution indicated that the subset of first three, or first seven, modes were separable.

4.3.5.3 Generating new conformations

Since each mode corresponds to a fluctuation between two oppositely directed motions, both directions being equally probable, two sets of deformations were considered for each mode, referred to as 'plus' or 'minus' displacements along the particular mode axis. The corresponding 'target' conformations, represented by the 3N-dimensional position vectors $\mathbf{R}^{+}_{k}$ and $\mathbf{R}^{-}_{k}$, are evaluated from

$$\mathbf{R}^{\pm}_{k} = \mathbf{R}^{0} \pm s\lambda_{k}^{-1/2}\,\mathbf{u}_{k}^{\mathrm{ANM}} \tag{4.36}$$

where $\mathbf{R}^{0}$ is the conformation prior to the application of restraints, and s is a scaling parameter, set by requiring the root-mean-square displacements (RMSD) in C^{α}-atom positions to remain close to 1.5 Å after reconfiguration along mode 1, and close to $1.5(\lambda_i/\lambda_1)^{1/2}$ Å for mode i. Note that $\boldsymbol{u}_{k}^{\mathrm{ANM}} = [\,\Delta\bar{R}_1(k)\;\Delta\bar{R}_2(k)\;\ldots\;\Delta\bar{R}_N(k)\,]^{\mathrm{T}}$. Each C^{α} atom is harmonically restrained to approach the target conformations, using a uniform spring constant of 1 kcal/(mol·Å^2) for all residues in ANM-restrained MD simulations. After reaching the two target structures $\mathbf{R}^{+}_{k}$ and $\mathbf{R}^{-}_{k}$, for a given mode, both are subjected to energy minimization (1000-3000 steps of steepest descent) without any restraints to relieve possible unrealistic distortions and to select the lower energy conformer among the two (see the profiles in Figure 4.6).

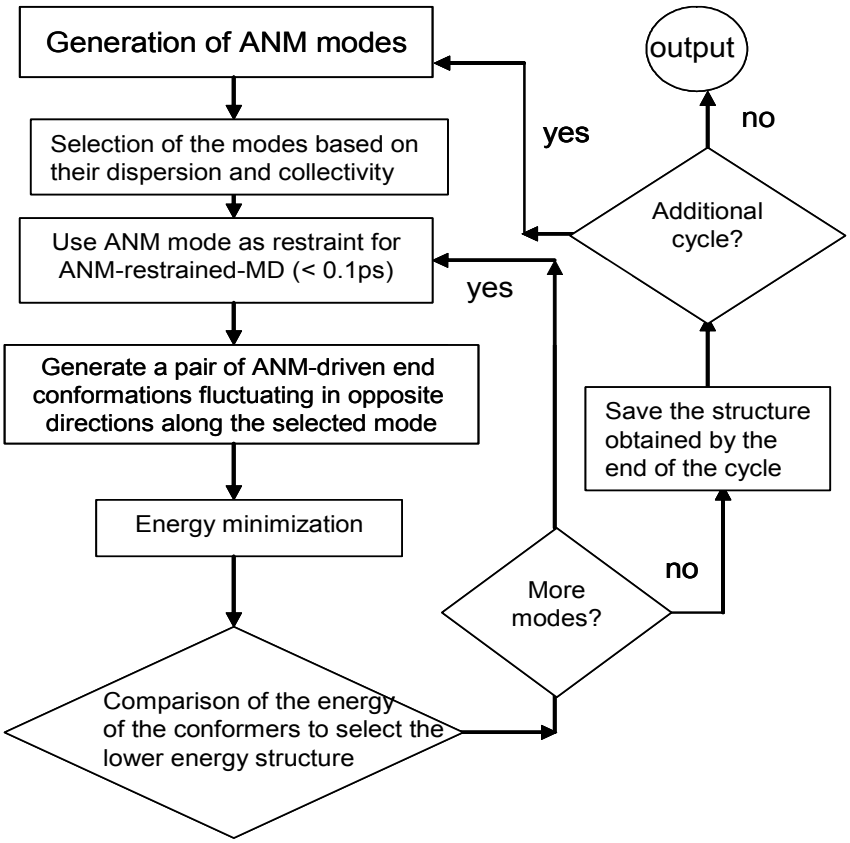

Figure 4.4 ANM-restrained MD protocol

4.3.6 The initial structure

Previously, Saam et al. (Saam et al., 2002) performed MD simulations to study the photoisomerization of retinal from 11-*cis* to all-*trans* configuration and the relaxation of rhodopsin succeeding this isomerization. The system consisted of rhodopsin embedded into a lipid bilayer with 108 POPC molecules, 6596 water molecules and 27 ions (Figure 4.5). The size of this system is 39,964 atoms. All titratable groups in the protein were considered to be charged, with the exception of Glu122 and Asp83 (Fahmy et al., 1993). After minimization and equilibration of the system, the isomerization of 11-*cis*-retinal into all-*trans*-retinal was induced by transiently switching the dihedral potential energy function of the corresponding bond. We

adopted the conformation at 495 ps of the MD simulations performed (ref) after this isomerization has been achieved, as our *initial* structure, because experimental evidence shows that rhodopsin undergoes the conformational changes to reach its active Meta II state only after the *cis-trans* isomerization of retinal and the accompanying flip of the *β*-ionone ring. We refer to the structure that is energy minimized and equilibrated in the presence of water and lipid molecules right before the isomerization of 11-*cis*-retinal, as the 'dark state'.

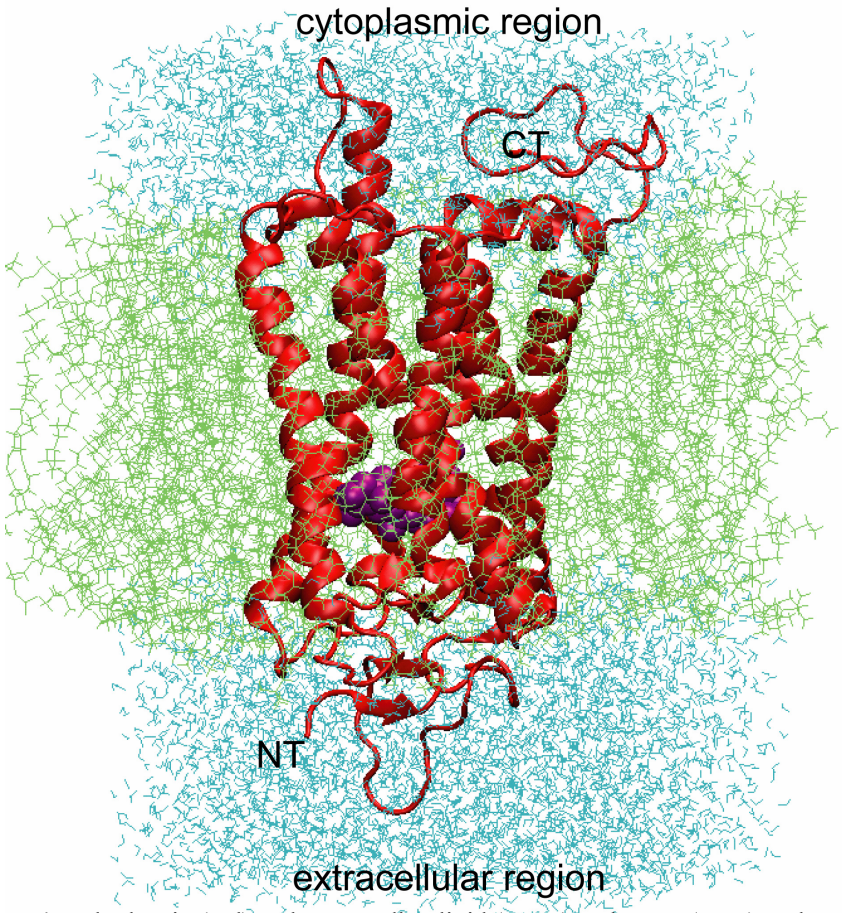

Figure 4.5 Rhodopsin (red) and surrounding lipid (green) and water (cyan) molecules

The results presented below were confirmed to be reproducible in two independent runs. In particular, the critical interactions in the chromophore binding pocket between the retinal atoms and neighboring helical residues, as well as the highly fluctuating sites on the H6 and CP region

that are known to be important for G-protein binding (Sakmar, 1998), were found to be robust features recurring in the simulated structural dynamics.

4.4 RESULTS AND DISCUSSION

4.4.1 Time evolution of RMSD from the initial state and accompanying change in energy

Figure 4.6 illustrates the time evolution of RMSD in C^{α}-positions from the initial structure (top panel), and the accompanying change in energy (lower panel) in a given cycle of our protocol. The two curves refer to the opposite direction deformations along each normal mode used as targets in MD simulations. Seven slowest modes are used in this cycle, each of them being succeeded by a short energy minimization algorithm. The conformer with the lower energy is chosen as the starting structure for the implementation of the successive modes. By the end of this cycle, the conformation departs from the original one by an RMSD of about 2 Å. Two such cycles were performed, to reach an RMSD of 3.5 Å including highly flexible CL2 and CL3, and two independent runs of two cycles each were performed to verify the reproducibility of the results.

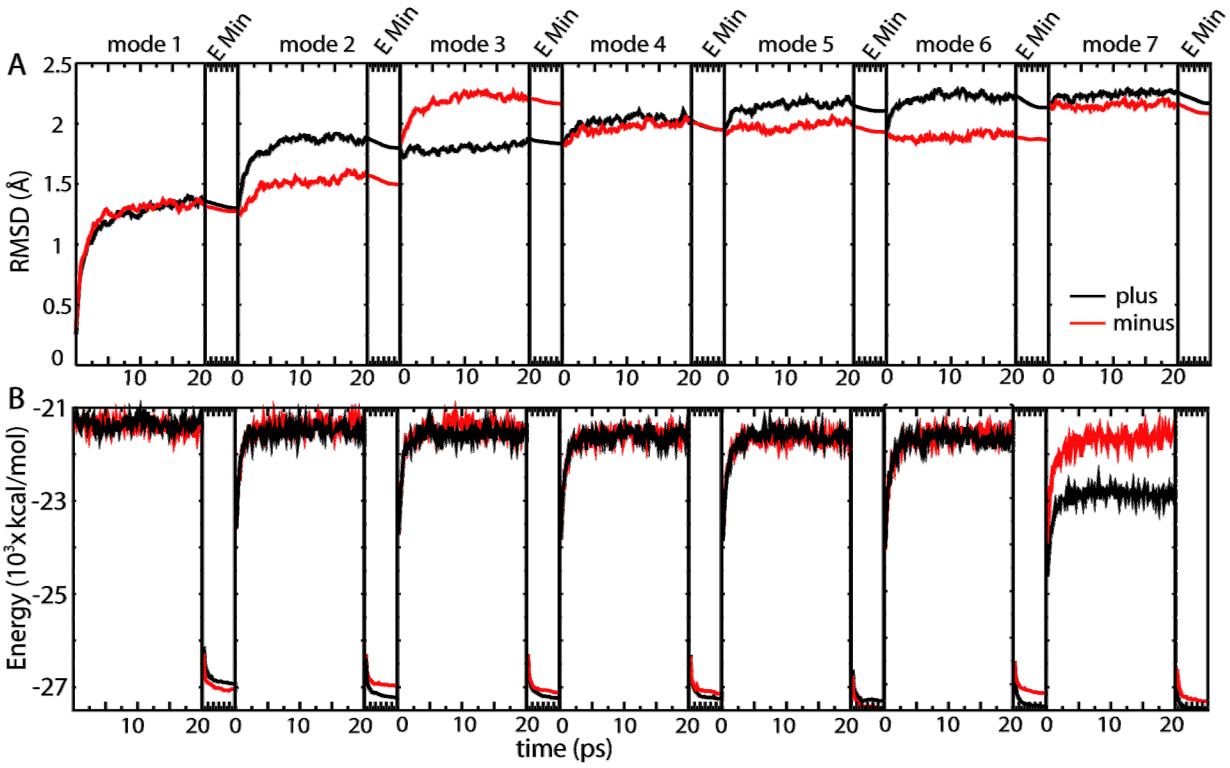

Figure 4.6 Time evolution of the overall RMSD in the α-carbon coordinates compared to the starting conformation (panel A) and its energy (panel B) for the first inner cycle of the ANM-restrained MD protocol
The two curves refer to the trajectories sampled with ANM restraints along opposite directions, each succeeded by a short energy minimization that selects the lower energy conformation for the next run (with another mode).

4.4.2 Two highly stable regions: the chromophore binding pocket and the CP ends of H1, H2 and H7

Figure 4.7A shows the RMSD profiles of residues from the initial structure obtained by the end of cycles 1 (red curve) and cycle 2 (black curve). The residue ranges of the TM helices H1-H8 are indicated by the gray bands. The ribbon diagram in panel B is color coded by the RMSDs observed by the end of the second cycle, from red (least mobile) to blue (most mobile). We see

that the two profiles in panel A share similar features. We are interested, in particular, in the minima (labeled) that are maintained by the end of the two cycles, which point to residues participating in highly stable hinge sites. Except for Pro23, all such residues are clustered in two regions on the 3D structure: the chromophore binding pocket and the CP end of H1, H2 and H7, which we will shortly refer to as chromophore-binding and CP regions. Interestingly, located separately from these two regions, one minimum is at Pro23 near the N-terminus. The Pro23His mutation is associated with the most frequently occurring form of autosomal dominant retinitis pigmentosa (ADRP), a hereditary progressive blinding disease. The pathogenicity of human mutant Pro23His such as photoreceptor degeneration was established by transgenic mice strain experiments (Olsson et al., 1992;Wang et al., 1997). It was suggested that an essential structural orientation between the N-terminus and EL1 was disrupted in the NT domain of Pro23His mutant.

4.4.3 The first highly stable (hinge) site near the chromophore directly participates in rhodopsin activation

This site includes Ala124 and Leu125 at H3; Trp161 and Cyc167 at H4; Pro180, Glu181, Ser186, Cys187 and Gly188 on the β-strands at EL2; Met207, His211, Phe212 at H5; Trp265, Tyr268 at H6; and Tyr292 and Ala295 at H7 (Figure 4.8). This cluster includes residues located in the retinal binding pocket at the dark state such as Trp265 and Tyr268, as well as residues that form new interactions to stabilize all-*trans*-retinal at the structure reached by the end of our simulations (e.g., Cys167; see below). Evidently, they would be very sensitive to conformational changes in retinal and would directly participate in rhodopsin activation.

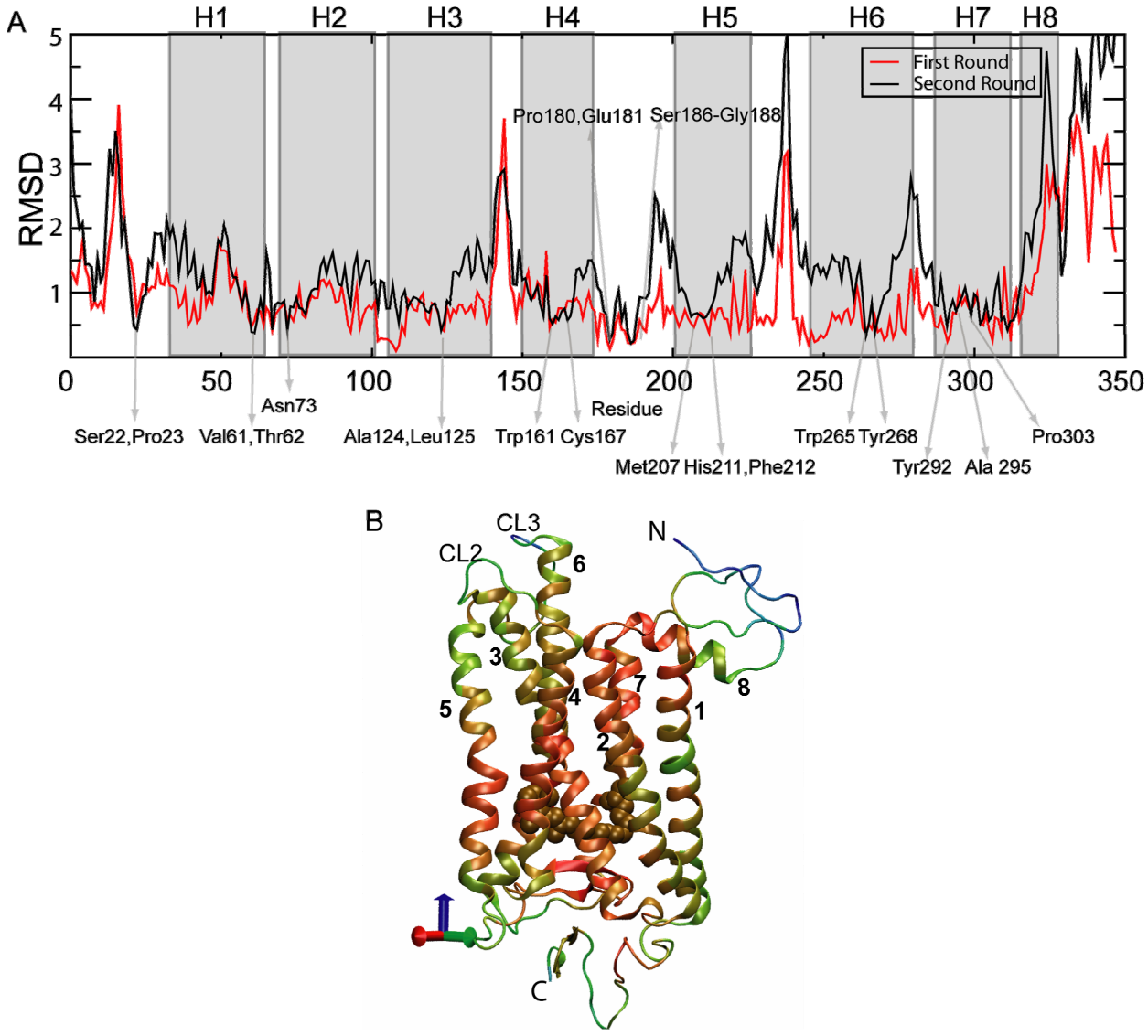

Figure 4.7 RMSD profile as a function of residue index, reached by the end of cycles 1 and 2 of the ANM-restrained MD protocol.

(A) The RMSDs exhibit an overall increase between cycles 1(red) and 2 (black), although the overall profile tends to maintain similar features. The sequence ranges of the helices (H1-H8) are indicated by the labels on the upper abscissa and distinguished by gray bands. Residues that lie at the minima are labeled. Non-TM regions exhibit higher mobilities in general, especially CL2 (between H3 and H4) and CL3 (between H5 and H6). (B) Ribbon diagram of rhodopsin color-coded according to the RMSD profile reached at the end of cycle 2. The color code is orange (smallest motions), yellow, green, blue (largest motions). The all-*trans* retinal is colored brown. Residues that exhibit the lowest RMSDs are clustered in two regions: around the chromophore, and in the CP portion of helices 1, 2, and 7. The CP ends of H3, H4, H5 and H6, including the loops CP2 and CP3, exhibit high RMSDs.

Two members of this cluster, Cys187 and Cys110, form a disulfide bridge. Mutation of both residues was also found to be associated with ADRP (Sakmar et al., 2002), most probably caused by the destabilization of the opsin structure near the chromophore binding site in both the Meta II from and dark state of rhodopsin. Cys187 forms the deepest minimum in the second cycle curve (black curve in Figure 4.7). Its RMSD and position does not change between the first and second cycles. Cys110 is in the deepest minimum of the first cycle curve. While still being one of the minima in the second, its RMSD is slightly higher than that of the first cycle.

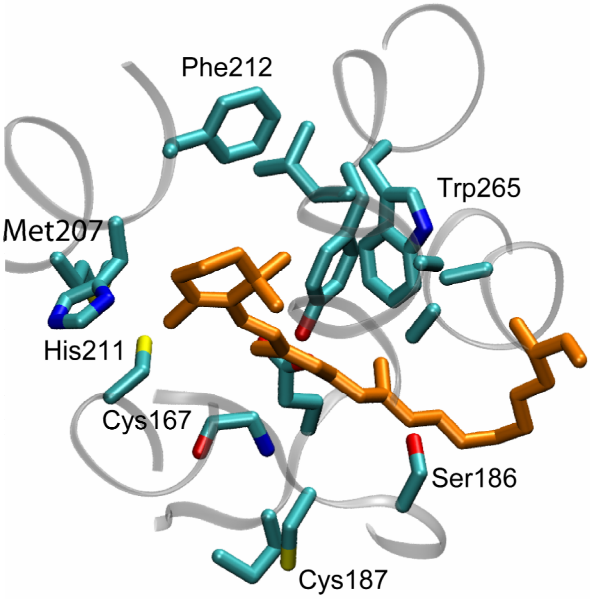

Figure 4.8 Hinge residues at the vicinity of the chromophore viewed from the CP regions. These are residues distinguished by their high stability (low mobility) in Figure 4A. They include Cys167 at H4, Ser186 and Cys187 on the β-sheet, Met207, His211 and Phe212 at H5, and Trp265 at H6. All-*trans*-retinal is shown in orange.

4.4.4 Two water molecules hydrogen-bonded to highly conserved residues in the cytoplasmic site

There is an extensive interhelical hydrogen-bond network between H1, H2 and H7. This network includes highly conserved residues in the GPCR family including the interhelical N-D pair (Asn55 at H1, Asp83 at H2). We also note that the NPXXY motif between Asn302 and Tyr306 at H7 also participates in this network of hydrogen bonds. These residues participate in the low RMSD cluster (CP region) that we have presently identified.

The interior of rhodopsin contains water molecules that mediate interhelical interactions, and it has been noted that these interactions could play a critical role in regulating the activity of GPCRs and the spectra sensitivity in visual pigment (Okada et al., 2002). Present simulation reveal that two residues belonging to the NPXXY motif at H7, Asn302 and Tyr306, are connected to H1 and H2 through water molecules located at the cavity between helices H1, H2 and H7. Notably, at the resulting conformation, Asn55 (H1), Asp83 (H2), and Asn302 (H7) are hydrogen bonded to a central water molecule (Figure 4.9A); and a second water molecule interacts closely with Thr62 (H1), Asn73 (H2), and Tyr306 (H7) (Figure 4.9B). This second water molecule was already found to be interacting with these residues by X-ray crystallography, and still remains at the same position despite the implementation of several cycles ANM-restraints in MD simulations. Overall ~ 20 water molecules are observed to span the helical bundle from the EC to the CP region and some exchange neighbors during the simulations. They mostly interact with H1, H2, H3, H4, and H7 residues.

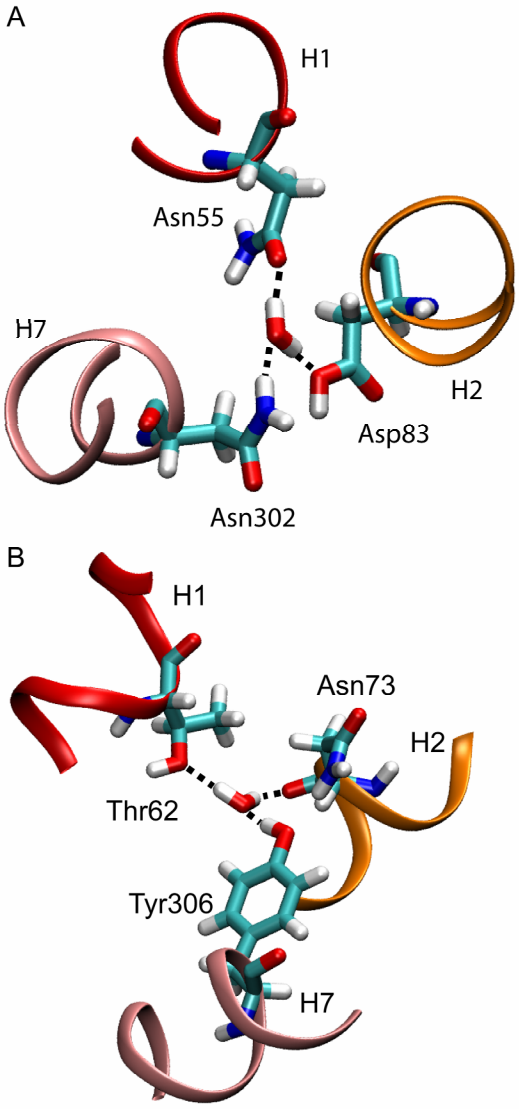

Figure 4.9 Water molecules that take part in the hydrogen bond network between H1, H2 and H7. Two water molecules are found be connected to highly conserved residues throughout the simulations, displayed in the respective panels A and B.

4.4.5 Comparison of the starting and ending structures

To visualize the conformational changes in the helices and the retinal, we have superimposed the initial and final structures in Figure 4.10A. The transparent and opaque ribbon diagrams correspond to the initial and final conformations, respectively. The largest conformational changes occur at the loops CL2, CL3 and CT end (see the analysis of the CP ends of rhodopsin below). In contrast, loops connecting the helices at the EC region exhibit lower displacements. Although the NT end does not have a secondary structure, the finger-like shape of the projecting loop is maintained throughout the simulations. The lower mobilities of the EC loops and NT end help maintaining the important interactions between the residues at these sites such as those involving Tyr102 at EL1 and Pro23 at NT end. Both in the initial and final structure, the retinal is in all-*trans* conformation since the initial conformation is taken from the MD simulation right after the isomerization of the 11^{th} bond on retinal and the flip of the β-ionone ring. Yet, at the resulting structure, the chromophore assumes a relatively more extended form (Figure 4.10B; shown in red).

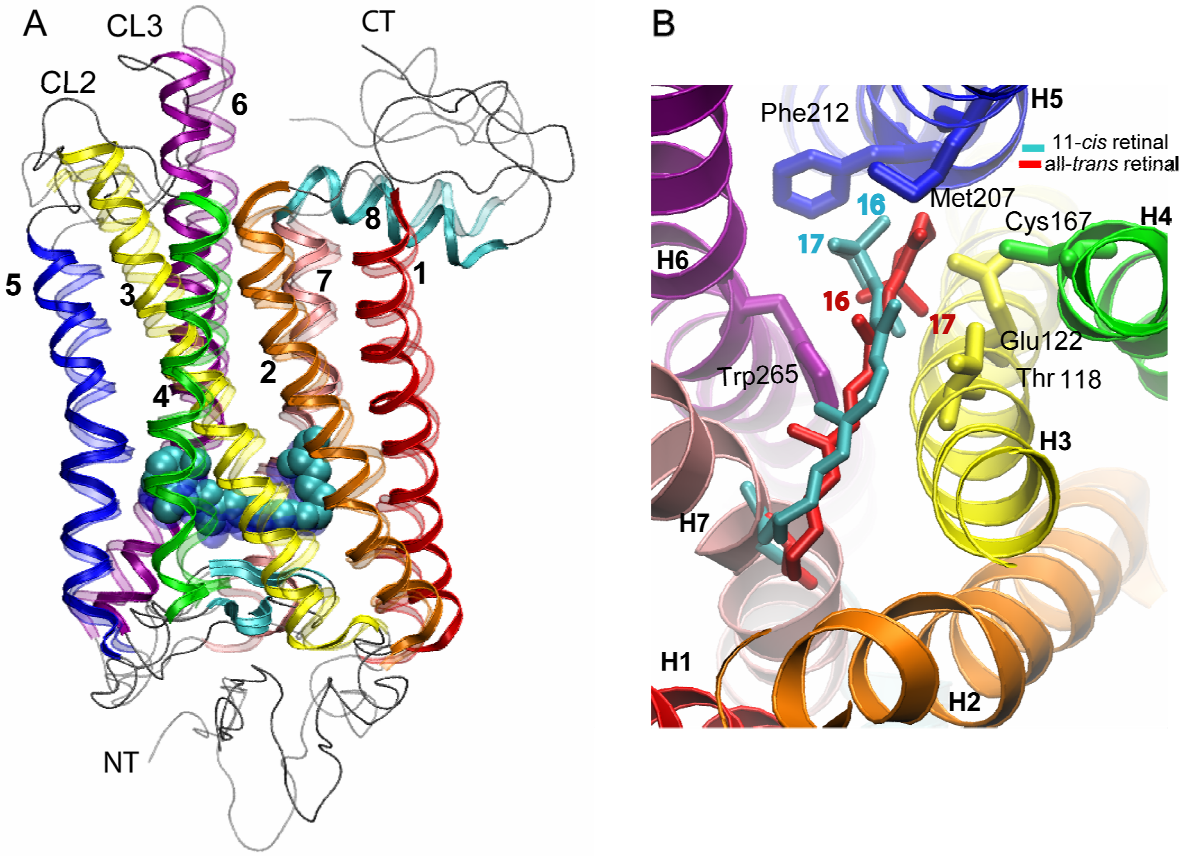

Figure 4.10 Comparison of the dark state (initial) with the proposed reconfigured state (final) and the reconfiguration of the retinal neighborhood.

(A) The resulting conformation (opaque) is superimposed on the initial conformation (transparent). (B) Redistribution of interactions in the neighborhood of the retinal. To clarify the change in the position of the β-ionone ring between the *cis* and *trans* conformations, atoms 16 and 17 are labeled in both conformations. Initially, 11-*cis* retinal was in close contact with H6 residues such as Phe261, Trp265, Tyr268 and Ala269. At the end of the simulation, the all-*trans*-retinal made new atomic contacts including Cys167 at H4, Phe203 at H5. The number of atom-atom contacts with Met207 and His211 increased in agreement with cross-linking and NMR experiments (Borhan et al., 2000, Patel et al., 2004). See the text for more details.

4.4.6 Rearrangements in the chromophore binding pocket to accommodate and stabilize the all-*trans*-retinal

The chromophore binding pocket is a densely packed region. Even small conformational changes in the retinal would be sufficient to significantly affect the interactions at this region, and violate the van der Waals volumes of interacting atoms. The isomerization of the retinal to *trans* state and the accompanying flip of the β-ionone ring indeed cause steric clashes between the all-*trans*-retinal and surrounding residues (Isin et al., 2006). These clashes are relieved upon rearrangements in the positions and orientations of helices, which in turn induces a redistribution of contacts in the chromophore binding pocket. We note in particular that the number of atom-atom contacts between Trp265 and the retinal in the *trans* form (based on interatomic interaction range of 4.5 Å) is significantly lower than those made in the *cis* form. The β-ionone ring of 11-*cis*-retinal was approximately parallel to the aromatic ring of Trp265, such that many atom-atom contacts were possible in the original structure. In contrast, only one atom, C16, on the *trans*-retinal is in the vicinity of Trp265 (Figure 4.10B). Furthermore, Phe261, Tyr268 and Ala269 at H6 make contacts with the β-ionone-ring to stabilize the 11-*cis*-retinal in dark state (not shown), while these contacts are lost in the new structure. Instead, a new amino acid, Cys167 at H4 lines the chromophore binding pocket and interacts with the atoms C3, C4, C5, C18 at the β-ionone ring of all-*trans*-retinal. Additionally, two residues at H5, Phe203 and His211, form new contacts with the atoms C18 and C2, C3, C4 of the retinal. The number of contacts made by Met207 and Thr118 is larger in the resulting structure compared to those in the dark state. Changes in contacts at the polyene chain mostly occur at C20. In the dark state, C20 interacts with Trp265, Tyr268 and Ala292. In the newly obtained structure, on the other hand, it interacts with Glu181, Ser 186, Cys187 on the β strands, and with Tyr268 (H6) and Ala295 (H7).

Some of these results are confirmed by cross-linking (Nakayama and Khorana, 1990;Borhan et al., 2000;Palczewski et al., 2000;Patel et al., 2004) and NMR experiments (Patel et al., 2004). In the dark state, 11-*cis* retinal cross-links to Trp265, while the all-*trans*-retinal cross-links to Ala169 instead of Trp265, consistent with the proximity redistribution of contacts observed here. Further investigation by high resolution solid state NMR measurements also showed that Trp126 and Trp265 interact more weakly with the retinal in the active state. Additionally, both the side chain of Glu122 and the backbone carbonyl of His211 are disrupted in Meta II (Patel et al., 2004).

4.4.7 CP ends of H3, H4, H5, H6 and connecting loops, CL2 and CL3, exhibit high mobility

Figure 4.11 shows the top view of rhodopsin (panel A), and the side view of helices close to the ERY motif (panel B). The resulting structure (opaque) is superposed on the dark state (transparent) to facilitate the comparison. The motions of helices are indicated by the arrows. The CP ends of H3, H4, H5 and H6, and the loops CL2 and CL3 that connect these helices at the CP region, are very mobile as evidenced by their high RMSD values, in agreement with spin labeling and cross-linking experiments (Nakayama and Khorana, 1990;Farrens et al., 1996;Cai et al., 1999;Borhan et al., 2000;Meng and Bourne, 2001;Hubbell et al., 2003). Arrangements of these helices lead to the exposure of the ERY motif at H3. The CP ends of H1, H2 and H7, on the other hand, closely maintain the configuration they assumed in the dark state structure (Figure 4.7A).

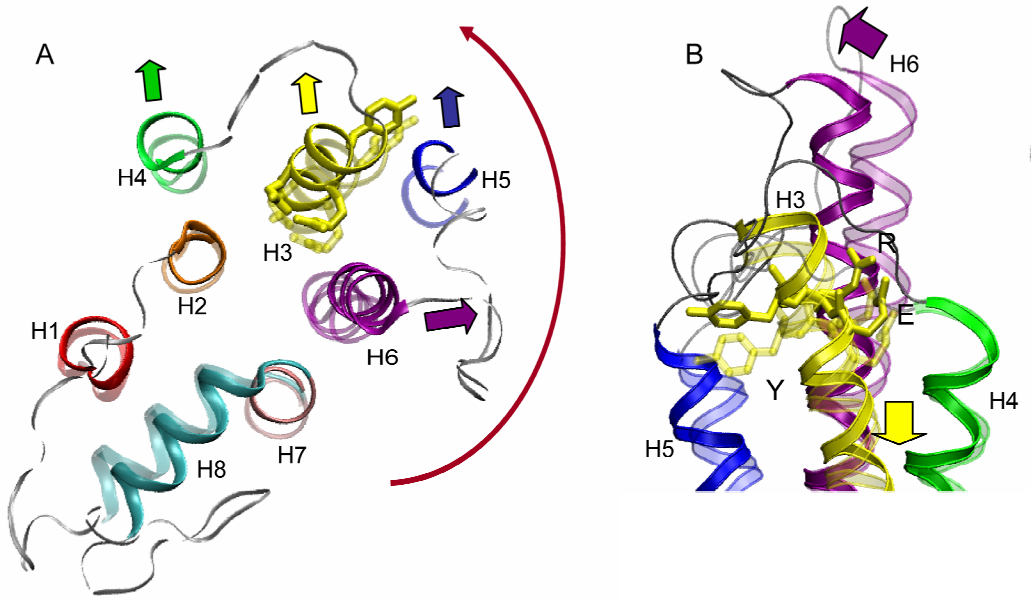

Figure 4.11 Comparison of the dark state (initial) and the resulting conformation of the CP end of the helical bundle (A) near the ERY motif (B) at H3.
The transparent and opaque ribbon diagrams correspond to the initial and final conformations, respectively. Arrows indicate the displacements of the helices.. Arrangements of H3, H4, H5, and H6 along with the loops CL2 and CL3, result in the exposure of the ERY motif at H3. H1, H2 and H7 undergo relatively small fluctuations.

4.4.8 Refinements in previous model inferred from present simulations

In the previous chapter, the side chain conformations were generated by assuming that they accompany the backbone motion, and then, to correct for possible unrealistic distortions in bond length and bond angles, the conformations were subjected to a short energy minimization in vacuum (Isin et al., 2006). In the present study, using ANM modes in SMD, we were able to generate conformations favored by a detailed atomic force field. Some residues identified in our previous work to act as global hinges maintained the same character here. These are all residues close to the chromophore: Ala124 and Ile125 on H3, Trp161 on H4, Phe212 on H5, and Pro180

and Cys187 on the β-sheet of the EC2. We also observe that the hinge site broadened in two directions, to include Cys167 on H4, Glu181, Ser186 on the β sheet of the EC2, Met207, His211 on H5, Trp265, Tyr268 on H6, and Tyr292, Ala295 on H7. Previously, we detected steric clashes between Cys167 and the C1, C2 and C16 atoms of all-*trans*-retinal in the dark state. Here, this residue is in close contact with the retinal and is positioned at the hinge region. The hinge residues at H1, H2 and H7 are relatively closer to the CP region where two water molecules are found to further stabilize the hinge site. Inclusion of explicit water in our model thus contributed to refine the precise location of the global hinge region, as well as identify a hydrogen bond network consolidated by water molecules. The conserved N-D pair (Asn55 and Asp83) also takes part in this network and is connected to Asn302 through a bridging water molecule.

4.5 CONCLUSION

4.5.1 ANM-Restrained MD method. Utility and limitations.

In the present thesis, we took advantage of the normal modes generated by the ANM to be able to sample global conformational changes during the MD simulations of rhodopsin. A major advantage of adopting such an ANM-restrained MD protocol is the ability to observe the cooperative changes at full atomic detail in the presence of explicit water and membrane molecules. To the best of our knowledge, this is the first TMD/SMD method that uses ANM modes. TMD requires selecting the target conformations, and cannot be used unless an end conformation is accessible. SMD also requires implementing a force, or acceleration, along a

predefined direction. In the present simulation the 'bias' used is not arbitrary, but that expected to be intrinsically preferred by the structure, as predicted by the ANM analysis previously verified to yield results consistent with experiments (Isin et al., 2006). We note that the simulations are performed at physiological temperatures, i.e. there is no need to activate transitions using high temperatures, which is probably another utility of adopting ANM restraints. Finally, ANM results are analytical. They do not necessitate the generation of a trajectory, in contrast to EDA, or optimal superimposition of appropriate domains before PCA (Hayward and Go, 1995;Hayward et al., 1997) and consequently, they are not affected by convergence or inadequate sampling problems that may affect PCA-based analyses of MD trajectories. Knowledge of one (equilibrium) structure is sufficient to predict the most likely (lowest energy ascent) directions of deformations (as the slowest modes) away from that local minimum. A limitation is however the lack of knowledge on the size of motions, which are now being defined by the force fields that control the MD simulations, and subsequent energy minimization steps (with no restrains imposed) where excessive, physically unrealistic deformations are being restored. Finally, this method is relatively simple enough to be easily implemented using the ANM modes in NAMD. We have recently developed a fully automated server that releases these modes for any PDB structure with known C^{α}-coordinates (or any model submitted as query in PDB format) (Eyal et al., 2006;Eyal et al., 2007). The incorporation of ANM restraints in MD packages such as NAMD (Kale et al., 1999) is straightforward and flexible. In future work, ANM-restrained MD can be implemented as an option in NAMD or relevant MD software packages.

4.5.2 Efficient propagation of signals from chromophore binding pocket to G-protein binding cytoplasmic loops

Although a wide range of information on both dark and active state of rhodopsin exists, the details of activation mechanism remains unclear. Here, we gained insights about the global motions that cooperatively propagate the local motions in the neighborhood of the isomerizing chromophore, to the G-protein binding sites on the CP surface. Two global hinge centers have been identified, which ensure this functional communication between distant positions. The first contains specific residues at H3, H4, H5, and H6 which directly sense the local conformational changes induced by retinal isomerization. Some of these residues stabilize 11-*cis*-retinal in the dark state (e.g., Trp265 and Tyr268 at H6) while others interact with the all-*trans*-retinal in the resulting structure (e.g., Cys167, Met207, and His211). In particular, the new contact found between Cys167 at H4 and the all-*trans*-retinal support our previous experiments showing the significant effect of Cys167Ser on Meta II decay rate (Isin et al., 2006). The second hinge, on the other hand, is located close to the CP end of helices H1, H2 and H7, and includes the NPXXY motif.

In the dark state structure, H6 is highly constrained by 11-*cis*-retinal. The aromatic ring of Trp265 makes several contacts with the β-ionone ring. Isomerization into *trans* form releases the constraints imposed at H6. In contrast, H4 and H5 form new contacts, and those made with H3 residues are redistributed. To accommodate these alterations in the chromophore binding site, H3, H4, H5 and H6 undergo conformational changes, which are reflected on their CP ends. Note that the local rearrangements of these helices near the chromophore binding packet efficiently propagate into these distal points due to the tight packing at the retinal-binding pocket, and the

internal rigidity (like a moment arm) of helices. The enhanced ability of secondary structural elements, including in particular helices, for efficiently transmitting signals to distant loci was pointed out in a recent study (Chennubhotla and Bahar, 2007) Interestingly, some residues in the chromophore binding area (e.g. disulfide bridge Cys110-Cys187) are particularly rigid as evidenced by their minimal RMSD by the end of our runs. The motions are even accentuated at the CP loops connecting the helices H3, H4, H5 and H6, leading to the exposure of ERY motif crucial for G-protein binding (Figure 4.7). Notably, in the activated structure, water molecules span the TM domain from the EC to the CP region. In addition to those detected by X-ray crystallography, new water molecules are located at the vicinity of highly conserved residues such as those in the cavity between Asn55, Asp83 and Asn302 (Figure 4.9A) .

5.0 CONCLUSIONS

5.1 USING LARGE SCALE MOTIONS FROM ELASTIC NETWORK MODELS TO GUIDE MD SIMULATIONS

Coarse-grained models used in NMA have been very useful in studying the collective dynamics of proteins and their complexes. Previous studies have demonstrated that the biologically relevant, long timescale motions of large structures can be predicted to a good approximation by the GNM and ANM, thus avoiding the need for expensive computations (Bahar et al., 1997;Atilgan et al., 2001;Chennubhotla et al., 2005;Bahar and Rader, 2005;Ma, 2005).

In this study, we first focused on the identification of the intrinsic dynamics of rhodopsin by the GNM and ANM. These analyses are based on the premises that (i) the topology of inter-residue contacts essentially determines the global motions of proteins, and (ii) the global modes are robust, i.e. they are uniquely defined by the equilibrium structure and are insensitive to the details of atomic interactions, as elaborated in several recent coarse-grained analyses of biomolecular equilibrium dynamics {Bahar, 2005 9147 /id;Chennubhotla, 2005 9146 /id;Ma, 2005 297 /id}. The major advantage of the GNM and ANM is the fact that the motions are determined from an *analytical* solution, i.e. the solution is deterministic for a given architecture; it is a unique function of the particular topology of contacts. It requires neither an energy minimization prior to calculations, nor the inclusion of any specific energy parameters, in

contrast to conventional NMA that employs detailed force fields and requires an energy minimization scheme to ascertain that the structure occupies an energy minimum.

Despite their insightful applications, ENM methods also have limitations. They lack information on residue specificities, atomic details or side chain motions. Furthermore, identifying which particular modes are involved in biologically relevant motions, and what their amplitudes are, are usually open questions. We also note that, although many functional mechanisms appear to be intrinsically defined by the 3D structure of the protein (Bahar and Rader, 2005;Eisenmesser et al., 2005;Changeux and Edelstein, 2005), the effects of the interactions with the environment such as the lipids and water molecules are yet to be assessed for membrane proteins. Such effects are usually accounted for by MD simulations. The latter suffer however from sampling inefficiencies and inability to explore time scales of biological interest.

Our approach to overcome these limitations has been to use the normal modes generated by the ANM as restraints in MD simulations, so as to sample cooperative and global conformational changes relevant to the biological function of rhodopsin. We developed the first ANM-restrained MD protocol which constrains the system minimally. It reduces the degrees of freedom of our system to those that are postulated to be important for the biological function. By this protocol, we observed cooperative changes with full atomic detail in the presence of the environment, i.e. explicit water and lipid molecules. TMD and SMD simulations require selecting a series of directions for the applied force. Rather than choosing the force directions randomly or guessing a direction of reconfiguration, as typically performed on the basis of

structural data, we used the modes of motions found by ANM analysis to be intrinsically favored by the overall architecture and proven to be functionally important.

This method allows us to observe the global motions that are beyond the reach of traditional MD simulations in a reasonable CPU time. The cost of a step in ANM-restrained MD protocol is almost the same as in unbiased MD simulations. The same method can also be used to apply a perturbation to only a certain portion of the biomolecular system and observe the response of the overall system. Furthermore, these simulations are performed at room temperatures by introducing ANM modes directly in MD simulations as restraints and there is no need to activate transitions using artificially high temperatures.

To extract meaningful information with regard to structural and dynamic properties such as the cooperative motions that involve multiple domain movements, ANM restrained-MD simulations can be performed without the necessity of generating a long trajectory for collective modes. This is in contrast to methods like EDA or PCA. ANM analysis permits us to avoid expensive computations by focusing on a small subset of low frequency normal modes that are extracted from the structure. Furthermore, conventional PCA analysis of MD trajectories may sometimes be inadequate for determining the interdomain/global motions due to the convergence problems or inaccuracy of MD data to which PCA is applied (Hayward and Go, 1995;Hayward et al., 1997).

5.2 THE ACTIVATION MECHANISM OF RHODOPSIN

The mechanism of rhodopsin activation consists of two main steps: (1) intrinsic conformational changes, triggered by retinal isomerization, which simultaneously (2) release structural constraints. First, we performed a detailed analysis of the GNM and ANM modes to characterize the global dynamics of rhodopsin in the dark state. Next, to identify the structural elements that stabilize rhodopsin, we simulated the thermal unfolding of rhodopsin using the FIRST software and analyzed the GNM high frequency modes of rhodopsin. Last, we used the newly developed ANM-restrained MD protocol to explore the global dynamics derived from multiple modes of ANM analysis of rhodopsin, while incorporating the effects from residue specificities in the presence of lipid and water molecules.

We compared our results to those from experiments performed, to gain insights into the molecular nature of the conformational changes accompanying activation. These experiments include extensive Cys scanning mutagenesis along the CP surface of rhodopsin in combination with site-directed spin labeling followed by EPR analysis of mobility, accessibility and spin-spin interactions, sulfhydryl reactivity, and disulfide cross-linking rates measurements. Furthermore, we tested the results by examining a set of Meta II fluorescence decay rates measured to empirically characterize the deactivation of rhodopsin mutants. We found that our model correctly predicts 93% of the experimentally observed qualitative response in 119 rhodopsin mutants for which the decay rates and misfolding data have been measured, including a systematic analysis of Cys->Ser replacements performed in Dr. Klein-Seetharaman's laboratory (Chennubhotla et al., 2005;Ma, 2005;Isin et al., 2006).

We observed that the transmission of the signal of retinal isomerization to the CP domain and induced conformational changes at the CP end to allow for G protein binding are ensured by a robust hinge site that does not change despite the application of several rounds of ANM restraints. This hinge is located in the close neighborhood the retinal binding pocket and consists of Ala124 and Leu125 on H3; Trp161 and Cys167 on H4; Pro180, Glu181, Ser186, Cys187 and Gly188 on the β-strands at EL2; Met207, His211, Phe212 at H5; Trp265, Tyr268 on H6; and Tyr292 and Ala295 on H7. It contains residues that are directly affected by the isomerization of retinal (e.g., Trp265 and Tyr268) and those stabilizing the resulting all-*trans* conformation of the retinal (e.g., Cys167 at H4, Met207, and His211 at H5). The rigidity of the disulfide bridge between Cys110 and Cys187 at the hinge region was also supported by FIRST calculations and the analysis of high frequency modes of GNM.

The chromophore binding pocket is a tightly packed region. The hinge site residues are sensitive to small conformational changes in retinal and directly participate in rhodopsin activation. The interactions at the hinge site with the retinal, predominantly with its ionone ring, restrict the motions of the transmembrane helix H6. In particular, the aromatic ring of Trp265 on H6 makes several contacts with the β-ionone ring of 11-*cis*-retinal. Isomerization of the retinal into the *trans* form and the flip of β-ionone ring release these interactions. We also observed that the dark state structure cannot accommodate the all-*trans*-retinal. The *cis*-*trans* isomerization potentially entails steric clashes between H4 and the β-ionone ring, which are relieved by the repositioning of H4. H4 and H5 form new contacts with all-*trans*-retinal. The global motions coordinated by the hinge region, the release of constraints or weakening of interactions between retinal and TMs, accompanied by the formation of new interactions are manifested by an

increased mobility towards the CP portions of H3-H6. This leads to the exposure of the ERY motif that is critical for G-protein binding and activation.

Interestingly, there is another stable region close to the CP ends of H1, H2 and H7, found by explicitly including water molecules in our ANM-restrained MD simulations. This Two water molecules are found form hydrogen bonds with a number of highly conserved residues of these helices. Two residues belonging to the NPXXY motif at H7, Asn302 and Tyr306, are connected to H1 and H2 through these two water molecules. Asn55 (H1), Asp83 (H2), and Asn302 (H7) are hydrogen bonded to a central water molecule; and a second water molecule bridges between Thr62 (H1), Asn73 (H2), and Tyr306 (H7). We also detected other water molecules that span the transmembrane domain from the EC to the CP region. Thus, in addition to those detected by X-ray crystallography, new water molecules were located at the vicinity of highly conserved residues.

In summary, there is a clear cooperativity between the changes in the chromophore environment and the conformational changes in the CP domain upon light-activation, which can be explained by the intrinsic dynamic properties of rhodopsin in the dark state. Our results support the view that the dark state structure is predisposed to undergoing the structural changes that are experimentally observed to be triggered upon light-induced retinal isomerization. This predisposition is revealed by the global vibrations (or relaxational fluctuations) naturally favored by the topology of inter-residue contacts before light-activation. Only a subset of the motions and interactions identified here has been studied to date experimentally. Our study thus provides a useful framework for experimental tests to validate the predictions and gain a better understanding of the molecular mechanisms/interactions that underlie rhodopsin activation.

6.0 FUTURE WORK

6.1 DETERMINING THE INTERACTION OF RHODOPSIN WITH TRANSDUCIN AND DEVELOPING A NUCLEOTIDE EXCHANGE MODEL FOR TRANSDUCIN

The molecular mechanism by which the binding of the active form of rhodopsin leads to the exchange of GDP to GTP from transducin is the least understood step in the visual transduction cascade. There is biophysical data for the interaction sites of rhodopsin and transducin, and several models have been developed toward explaining the mechanism of Meta II catalyzed transducin activation (Ernst et al., 1995;Acharya and Karnik, 1996;Bourne, 1997;Abadji et al., 1999;Ernst et al., 2000;Marin et al., 2000;Marin et al., 2001;Abdulaev et al., 2005;Palczewski, 2006;Abdulaev et al., 2006;Downs et al., 2006). However, how the binding of Meta II to transducin causes the release of GDP at a remote site is still an unanswered question and the mechanism of transducin activation has not been understood. The detailed structure of the rhodopsin-transducin complex also remains to be determined. By utilizing the information on rhodopsin activation obtained in this thesis, the next step would be to explore the interaction of rhodopsin with transducin and elucidate the functional motions of transducin for the activation and nucleotide exchange.

Currently, crystal structures of transducin at different conformations are available. These conformations include the heterotrimeric $G_{\alpha\beta\gamma}$ with GDP, the free $G_{\beta\gamma}$ dimer and the G_{α} subunit

complexed with the GTPγS, the GDP and the GDP/wAIF$^-_4$ respectively (NOEL et al., 1993;Lambright et al., 1994;Lambright et al., 1996). These structures provide static snapshots of transducin at certain stages of the visual transduction cascade. A comparative examination of these structures would be the first step for understanding the conformational changes necessary for the activation and nucleotide exchange of transducin.

Meta II specifically interacts with the heterotrimeric form of transducin, $G_{\alpha\beta\gamma}$. The C-terminal end of G_α is shown to be the receptor recognition region, and a cluster of residues belonging to G_α are critical for the nucleotide exchange (Abdulaev et al., 2005;Abdulaev et al., 2006). Although the role that $G_{\beta\gamma}$ plays in GDP exchange is not as clear as G_α, it is determined that the C-terminus of G_γ also binds to Meta II. It is also suggested that $G_{\beta\gamma}$ has a mechanistic role in the exchange of GDP to GTP from $G_{\alpha\beta\gamma}$ (Sakmar et al., 2002;Abdulaev et al., 2006). To elucidate the mechanism of rhodopsin-catalyzed nucleotide exchange of transducin, it is necessary to analyze the structure of the heterotrimeric form, $G_{\alpha\beta\gamma}$.

The intrinsic motions of $G_{\alpha\beta\gamma}$ in GDP-bound form can be found by the GNM and ANM analysis, using for example the newly developed servers (Eyal et al., 2006). These motions would then be combined with the existing experimental data to develop a nucleotide exchange model for transducin. Previous comparisons of the dynamics of the complexed and free forms of different proteins have given insights about the effect of ligand binding on collective dynamics (Bahar and Jernigan, 1998;Yang and Bahar, 2005). Conformational motions obtained by the GNM and ANM could be compared to the existing structures of transducin at different stages to

find the correlations between the intrinsic dynamica and experimentally observed changes in structures.

6.2 EXTENDING THE KNOWLEDGE ON RHODOPSIN ACTIVATION MECHANISM TO OTHER GPCRS.

The crystal structure of rhodopsin has shown that rhodopsin is a prototypical member of the GPCR family A, and represents 90 % of all GPCR family members (Palczewski, 2006). Hence, the knowledge on the functional motions of rhodopsin gained in this work could be extended to other members of the GPCR family. Previous experiments suggested that most GPCR family members undergo similar functional motions (Hogger et al., 1995;Ballesteros et al., 1998;Horn et al., 1998;Ballesteros et al., 2001;Shi et al., 2002). We observed that the global motions of rhodopsin are not restricted to only the CP end of the molecule but there are cooperative motions of the TM helices including their EC portions. These motions can be important for GPCRs and may be evolutionarily conserved throughout the GPCR family. In addition, the robust hinge site around the chromophore area is observed to contain highly conserved residues. These residues may also be important for ligand binding of other GPCRs.

A more detailed study could be performed to investigate the role of hinge residues throughout the GPCR families. Ranganathan and coworkers developed a method to understand how signals originating at one site in a protein propagate to affect distant functional sites. Their method is based on statistical analysis of large number of sequences to map the global network of amino acid interactions in a protein (Suel et al., 2003). This method could be used to

understand the role of the hinge site residues in the propagation of the activation signal to other sites in rhodopsin. Furthermore, how this hinge site residues were evolutionarily altered among GPCRs to bind different ligands and how the remote G-protein binding sites are affected by this alteration could also be explored.

After the emergence of the crystal structure of the rhodopsin, structural models of many GPCRs have been determined by homology-based or comparative modeling. GPCRDB maintains a website for the 3D structures of GPCRs predicted by homology modeling (http://www.gpcr.org/models/index.html). Additionally, an elaborate method has been developed for prediction of 3D structures of GPCRs, namely Membstruk. Their method is an alternative to homology modeling and also be used to predict the structures of the GPCRs that have lower sequence similarity to rhodopsin. It consists of these essential steps: (i) the hydrophobicity analysis of the sequence to find the TM regions. (ii) the assembly of the helical bundle by using the low resolution electron density map of rhodopsin (Unger and Schertler, 1995). (iii) Monte Carlo optimization of the rotational orientation of the helices. (iv) loop building by WHATIF (Vriend, 1990). (v) the optimization of the final model in the presence of lipids by conjugate gradient minimization. The Membstruk's accuracy was tested by predicting the structures of rhodopsin and bacteriorhodopsin and validating them by comparing with the crystal structures of these proteins (Vaidehi et al., 2002). This strategy has been applied to a number of GPCRs and further validated by experimental data (Hall et al., 2004;Peng et al., 2006;Vaidehi et al., 2006). Among the structures that are predicted theoretically, β1-adrenergic receptor is one of the most extensively studied the GPCR family member and it also belongs to the same family of GPCRs as rhodopsin (family A) (Müller, 2000;Meng and Bourne, 2001;Ballesteros et al., 2001;Gouldson

et al., 2004). Therefore, it could serve as a suitable candidate for structural study as the next GPCR.

The structure of β1-adrenergic receptor determined by Membstruk is available and could be compared to those obtained by homology modeling, so as to view the differences. A pairwise sequence alignment with β1-adrenergic receptor and rhodopsin could also be performed to define regions that have sequence similarity. Then, the structure of β1-adrenergic receptor will be analyzed by the GNM and ANM. The slow modes of the GNM and ANM would be examined to see which global motions are common in both rhodopsin and β1-adrenergic receptor for the function of these proteins. Then, the global motions of β1-adrenergic receptor could be used in ANM-restrained MD to examine whether the robust hinge site that was preserved in rhodopsin, after several rounds restraining, also exists in β1-adrenergic receptor.

Another GPCR which is an evolutionary distant family member to rhodopsin would be could be examined. The structure of this GPCR could not be predicted accurately by sequence homology due to its low sequence homology to rhodopsin. Therefore, it would be modeled by both Membstruk and homology modeling. Then, these two models obtained by two different methods could be compared to determine the level of agreement between the two approaches. Then, if there is a reasonable agreement, the GNM and ANM analysis of the model could be performed to see which motions are conserved and which motions are different in this distant member of the family.

BIBLIOGRAPHY

Abadji, V., J. M. Lucas-Lenard, C. Chin, and D. A. Kendall. 1999. Involvement of the carboxyl terminus of the third intracellular loop of the cannabinoid CB1 receptor in constitutive activation of Gs. *J. Neurochem.* 72:2032-8.

Abdulaev, N. G., T. Ngo, E. Ramon, D. M. Brabazon, J. P. Marino, and K. D. Ridge. 2006. The receptor-bound "empty pocket" state of the heterotrimeric G-protein α-subunit is conformationally dynamic. *Biochemistry* 45:12986-12997.

Abdulaev, N. G., T. Ngo, C. Zhang, A. Dinh, D. M. Brabazon, K. D. Ridge, and J. P. Marino. 2005. Heterotrimeric G-protein α-subunit adopts a "preactivated" conformation when associated with betagamma-subunits. *J. Biol. Chem.* 280:38071-38080.

Abseher, R. and M. Nilges. 2000. Efficient sampling in collective coordinate space. *Proteins* 39:82-88.

Acharya, S. and S. S. Karnik. 1996. Modulation of GDP release from transducin by the conserved Glu134-Arg135 sequence in rhodopsin. *J. Biol. Chem.* 271:25406-25411.

Adcock, S. A. and J. A. McCammon. 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106:1589-1615.

Alewijnse, A. E., H. Timmerman, E. H. Jacobs, M. J. Smit, E. Roovers, S. Cotecchia, and R. Leurs. 2000. The effect of mutations in the DRY motif on the constitutive activity and structural instability of the histamine H(2) receptor. *Mol. Pharmacol.* 57:890-8.

Allen, M. P. and D. J. Tildesley. 2006. Computer Simulations of Liquids. Oxford Science Publications.

Altenbach, C., K. Cai, H. G. Khorana, and W. L. Hubbell. 1999a. Structural features and light-dependent changes in the sequence 306-322 extending from helix VII to the palmitoylation sites in rhodopsin: a site-directed spin-labeling study. *Biochemistry* 38:7931-7.

Altenbach, C., J. Klein-Seetharaman, J. Hwa, H. G. Khorana, and W. L. Hubbell. 1999b. Structural features and light-dependent changes in the sequence 59-75 connecting helices I and II in rhodopsin: a site-directed spin-labeling study. *Biochemistry* 38:7945-9.

Altenbach, C., K. Yang, D. L. Farrens, Z. T. Farahbakhsh, H. G. Khorana, and W. L. Hubbell. 1996. Structural features and light-dependent changes in the cytoplasmic interhelical E-F loop region of rhodopsin: a site-directed spin-labeling study. *Biochemistry* 35:12470-8.

Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. *Proteins* 17:412-425.

Amadei, A., A. B. Linssen, B. L. de Groot, D. M. van Aalten, and H. J. Berendsen. 1996. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* 13:615-625.

Anukanth, A. and H. G. Khorana. 1994. Structure and function in rhodopsin. Requirements of a specific structure for the intradiscal domain. *J. Biol. Chem.* 269:19738-44.

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505-515.

Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational Dynamics of folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys. Rev. Lett.* 80:2733-6.

Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single- parameter harmonic potential. *Fold. Des.* 2:173-181.

Bahar, I. and R. L. Jernigan. 1998. Vibrational dynamics of transfer RNAs. Comparison of the free and enzyme-bound forms. *J. Mol. Biol.* 281:871-884.

Bahar, I. and A. J. Rader. 2005. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15:586-592.

Ballesteros, J., S. Kitanovic, F. Guarnieri, P. Davies, B. J. Fromme, K. Konvicka, L. Chi, R. P. Millar, J. S. Davidson, H. Weinstein, and S. C. Sealfon. 1998. Functional microdomains in G-protein-coupled receptors. The conserved arginine-cage motif in the gonadotropin-releasing hormone receptor. *J. Biol. Chem.* 273:10445-53.

Ballesteros, J. A., A. D. Jensen, G. Liapakis, S. G. Rasmussen, L. Shi, U. Gether, and J. A. Javitch. 2001. Activation of the b_2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* 276:29171-7.

Baranski, T. J., P. Herzmark, O. Lichtarge, B. O. Gerber, J. Trueheart, E. C. Meng, T. Iiri, S. P. Sheikh, and H. R. Bourne. 1999. C5a receptor activation. Genetic identification of critical residues in four transmembrane helices. *J. Biol. Chem.* 274:15757-65.

Becker, O. M., A. D. J. MacKerell, B. Roux, and M. Wanatabe. 2001. Computaional Biochemistry and Biophysics. Marcel Dekker.

Berendsen, H. J. and S. Hayward. 2000. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* 10:165-169.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235-42.

Berson, E. L. 1993. *Retinitis Pigmentosa. The Friedenwald Lecture.* Invest. *Invest. Opthalmol. Vis. Sci.* 34:1659-76.

Borhan, B., M. L. Souto, H. Imai, Y. Shichida, and K. Nakanishi. 2000. Movement of retinal along the visual transduction path. *Science* 288:2209-2212.

Botelho, A. V., N. J. Gibson, R. L. Thurmond, Y. Wang, and M. F. Brown. 2002. Conformational energetics of rhodopsin modulated by nonlamellar-forming lipids. *Biochemistry* 41:6354-6368.

Bourne, H. R. 1997. How receptors talk to trimeric G proteins. *Curr. Opin. Cell Biol.* 9:134-142.

Brown, M. F. 1994. Modulation of rhodopsin function by properties of the membrane bilayer. *Chem. Phys. Lipids* 73:159-180.

Cai, K., J. Klein-Seetharaman, J. Hwa, W. L. Hubbell, and H. G. Khorana. 1999. Structure and function in rhodopsin: effects of disulfide cross-links in the cytoplasmic face of rhodopsin on transducin activation and phosphorylation by rhodopsin kinase. *Biochemistry* 38:12893-8.

Cai, K., R. Langen, W. L. Hubbell, and H. G. Khorana. 1997. Structure and function in rhodopsin: topology of the C-terminal polypeptide chain in relation to the cytoplasmic loops. *Proc. Natl. Acad. Sci. U. S. A* 94:14267-72.

Case, D. A., T. E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668-1688.

Changeux, J. P. and S. J. Edelstein. 2005. Allosteric mechanisms of signal transduction. *Science* 308:1424-1428.

Cheatham, T. E., J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 1995. Molecular-Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of Dna, Rna, and Proteins. *J. Am. Chem. Soc.* 117:4193-4194.

Chennubhotla, C. and I. Bahar. 2007. Markov Propagation of Signals in Proteins and its Relation to Equilibrium Fluctuations. *PLoS Comput. Biol.* in press.

Chennubhotla, C., A. J. Rader, L. W. Yang, and I. Bahar. 2005. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* 2:S173-S180.

Choi, G., J. Landin, J. F. Galan, R. R. Birge, A. D. Albert, and P. L. Yeagle. 2002. Structural studies of metarhodopsin II, the activated form of the G-protein coupled receptor, rhodopsin. *Biochemistry* 41:7318-7324.

Cohen, G. B., T. Yang, P. R. Robinson, and D. D. Oprian. 1993. Constitutive activation of opsin: influence of charge at position 134 and size at position 296. *Biochemistry* 32:6111-5.

Columbus, L. and W. L. Hubbell. 2002. A new spin on protein dynamics. *Trends Biochem. Sci.* 27:288-95.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 117:5179-5197.

Crozier, P. S., M. J. Stevens, L. R. Forrest, and T. B. Woolf. 2003. Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: coupling between local retinal and larger scale conformational change. *J. Mol. Biol.* 333:493-514.

Crozier, P. S., M. J. Stevens, and T. B. Woolf. 2007. How a small change in retinal leads to G-protein activation: initial events suggested by molecular dynamics calculations. *Proteins* 66:559-574.

Cui, Q. and I. Bahar. 2006. Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems. CRC Press, Taylor & Francis Group.

de Groot, B. L., A. Amadei, R. M. Scheek, N. A. van Nuland, and H. J. Berendsen. 1996. An extended sampling of the configurational space of HPr from E. coli. *Proteins* 26:314-322.

Demirel, M. C., A. R. Atilgan, R. L. Jernigan, B. Erman, and I. Bahar. 1998. Identification of kinetically hot residues in proteins. *Protein Sci.* 7:2522-2532.

Downs, M. A., R. Arimoto, G. R. Marshall, and O. G. Kisselev. 2006. G-protein α and βγ subunits interact with conformationally distinct signaling states of rhodopsin. *Vision Res.* 46:4442-4448.

Dryja, T. P. and E. L. Berson. 1995. Retinitis Pigmentosa and Allied Diseases. Implications of Genetic Hetergeneity. *Invest. Opthalmol. Vis. Sci.* 36:1197-1200.

Eisenmesser, E. Z., O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern. 2005. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438:117-121.

Ernst, O. P., K. P. Hofmann, and T. P. Sakmar. 1995. Characterization of rhodopsin mutants that bind transducin but fail to induce GTP nucleotide uptake. Classification of mutant pigments by fluorescence, nucleotide release, and flash-induced light-scattering assays. *J. Biol. Chem.* 270:10580-10586.

Ernst, O. P., C. K. Meyer, E. P. Marin, P. Henklein, W. Y. Fu, T. P. Sakmar, and K. P. Hofmann. 2000. Mutation of the fourth cytoplasmic loop of rhodopsin affects binding of transducin and peptides derived from the carboxyl- terminal sequences of transducin α and γ subunits. *J. Biol. Chem.* 275:1937-1943.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 103:8577-8593.

Eyal, E., C. Chennubhotla, L. W. Yang, and I. Bahar. 2007. Anisotropic Fluctuations of Amino Acids in Protein Structures: Insights from X-Ray Crystallography and Elastic Network Models. *Bioinformatics.* in press..

Eyal, E., L. W. Yang, and I. Bahar. 2006. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics.* 22:2619-2627.

Fahmy, K., F. Jager, M. Beck, T. A. Zvyaga, T. P. Sakmar, and F. Siebert. 1993. Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin II: a Fourier-transform infrared spectroscopy study of site-directed mutants. *Proc Natl Acad Sci U. S. A* 90:10206-10210.

Farahbakhsh, Z. T., K. D. Ridge, H. G. Khorana, and W. L. Hubbell. 1995. Mapping light-dependent structural changes in the cytoplasmic loop connecting helices C and D in rhodopsin: a site-directed spin labeling study. *Biochemistry* 34:8812-9.

Farrens, D. L., C. Altenbach, K. Yang, W. L. Hubbell, and H. G. Khorana. 1996. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* 274:768-770.

Farrens, D. L. and H. G. Khorana. 1995. Structure and function in rhodopsin. Measurement of the rate of metarhodopsin II decay by fluorescence spectroscopy. *J Biol Chem* 270:5073-6.

Finkelstein, A. V. and O. B. Ptitsyn. 2002. Protein Physics: A Course of Lectures. Academic Press.

Garriga, P., X. Liu, and H. G. Khorana. 1996. Structure and function in rhodopsin: correct folding and misfolding in point mutants at and in proximity to the site of the retinitis pigmentosa mutation Leu-125-->Arg in the transmembrane helix C. *Proc. Natl. Acad. Sci. U. S. A.* 93:4560-4.

Gouldson, P. R., N. J. Kidley, R. P. Bywater, G. Psaroudakis, H. D. Brooks, C. Diaz, D. Shire, and C. A. Reynolds. 2004. Toward the active conformations of rhodopsin and the beta2-adrenergic receptor. *Proteins* 56:67-84.

Greasley, P. J., F. Fanelli, O. Rossier, L. Abuin, and S. Cotecchia. 2002. Mutagenesis and modelling of the α($_1$β)-adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation. *Mol. Pharmacol.* 61:1025-32.

Grossfield, A., S. E. Feller, and M. C. Pitman. 2006. A role for direct interactions in the modulation of rhodopsin by omega-3 polyunsaturated lipids. *Proc. Natl. Acad. Sci. U. S. A* 103:4888-4893.

Grossfield, A., S. E. Feller, and M. C. Pitman. 2007. Convergence of molecular dynamics simulations of membrane proteins. *Proteins* 67:31-40.
Hall, S. E., W. B. Floriano, N. Vaidehi, and W. A. Goddard, III. 2004. Predicted 3-D structures for mouse I7 and rat I7 olfactory receptors and comparison of predicted odor recognition profiles with experiment. *Chem. Senses* 29:595-616.
Hayward, S. and N. Go. 1995. Collective Variable Description of Native Protein Dynamics. *Annu. Rev. Phys. Chem* 46:223-250.
Hayward, S., A. Kitao, and H. J. Berendsen. 1997. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins* 27:425-437.
Hespenheide, B. M., A. J. Rader, M. F. Thorpe, and L. A. Kuhn. 2002. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Graph. Model.* 21:195-207.
Hogger, P., M. S. Shockley, J. Lameh, and W. Sadee. 1995. Activating and inactivating mutations in N- and C-terminal i3 loop junctions of muscarinic acetylcholine Hm1 receptors. *J. Biol. Chem.* 270:7405-10.
Horn, F., R. Bywater, G. Krause, W. Kuipers, L. Oliveira, A. C. Paivia, C. Sander, and G. Vriend. 1998. The interaction of class B G protein-coupled receptors with their hormones. *Receptor Channels* 5:305-14.
Hubbell, W. L., C. Altenbach, C. M. Hubbell, and H. G. Khorana. 2003. Rhodopsin structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Membrane Proteins* 63:243-290.
Huber, T., A. V. Botelho, K. Beyer, and M. F. Brown. 2004. Membrane model for the G-protein-coupled receptor rhodopsin: Hydrophobic interface and dynamical structure. *Biophys. J.* 86:2078-2100.
Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Biol.* 14:33-8.
Hwa, J., P. Garriga, X. Liu, and H. G. Khorana. 1997. Structure and function in rhodopsin: packing of the helices in the transmembrane domain and folding to a tertiary structure in the intradiscal domain are coupled. *Proc. Natl. Acad. Sci. U. S. A.* 94:10571-6.
Hwa, J., J. Klein-Seetharaman, and H. G. Khorana. 2001. Structure and function in rhodopsin: Mass spectrometric identification of the abnormal intradiscal disulfide bond in misfolded retinitis pigmentosa mutants. *Proc. Natl. Acad. Sci. U. S. A.* 98:4872-6.
Hwa, J., P. J. Reeves, J. Klein-Seetharaman, F. Davidson, and H. G. Khorana. 1999. Structure and function in rhodopsin: further elucidation of the role of the intradiscal cysteines, Cys-110, -185, and -187, in rhodopsin folding and function. *Proc. Natl. Acad. Sci. U. S. A.* 96:1932-5.
Inglehearn, C. F., T. J. Keen, R. Bashir, M. Jay, F. Fitzke, A. C. Bird, A. Crombie, and S. Bhattacharya. 1992. A completed screen for mutations of the rhodopsin gene in a panel of patients with autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* 1:41-5.
Ishiguro, M., T. Hirano, and Y. Oyama. 2003. Modelling of photointermediates suggests a mechanism of the flip of the β-ionone moiety of the retinylidene chromophore in the rhodopsin photocascade. *Chembiochem.* 4:228-231.
Ishiguro, M., Y. Oyama, and T. Hirano. 2004. Structural models of the photointermediates in the rhodopsin photocascade, lumirhodopsin, metarhodopsin I, and metarhodopsin II. *Chembiochem.* 5:298-310.

Isin, B., P. Doruker, and I. Bahar. 2002. Functional motions of influenza virus hemagglutinin: a structure-based analytical approach
1. *Biophys. J.* 82:569-581.
Isin, B., A. J. Rader, H. K. Dhiman, J. Klein-Seetharaman, and I. Bahar. 2006. Predisposition of the dark state of rhodopsin to functional changes in structure. *Proteins* 65:970-983.
Isin, B., K. Schulten, E. Tajkhorshid, and I. Bahar. 2008. Mechanism of Signal Propagation upon Isomerization: Insights from Molecular Dynamics Simulations of Rhodopsin Restrained by Normal Modes. *Biophys. J.*
Israelewitz, B., J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten. 2001. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* 19:13-25.
Jacobs, D. J., A. J. Rader, M. F. Thorpe, and L. A. Kuhn. 2001. Protein Flexibility Predictions Using Graph Theory. *Proteins-Struct. Funct. Genet.* 44:150-165.
Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283-312.
Karnik, S. S., T. P. Sakmar, H.-B. Chen, and H. G. Khorana. 1988. Cysteine Residues 110 and 187 Are Essential for the Formation of Correct Structure in Bovine Rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* 85:8459-63.
Kaushal, S., K. D. Ridge, and H. G. Khorana. 1994. Structure and function in rhodopsin: the role of asparagine-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* 91:4024-28.
Keskin, O., S. R. Durrell, I. Bahar, R. L. Jernigan, and D. G. Covell. 2002. Relating molecular flexibility to function. A case study of tubulin
8988. *Biophys. J.* 83:663-680.
Khorana, H. G. 2000. Molecular biology of light transduction by the mammalian photoreceptor, rhodopisn. *J. Biomol. Structure and Dynamics* 11:1-16.
Kitao, A. and N. Go. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* 9:164-169.
Klco, J. M., C. B. Wiegand, K. Narzinski, and T. J. Baranski. 2005. Essential role for the second extracellular loop in C5a receptor activation. *Nat Struct. Mol Biol* 12:320-326.
Klein-Seetharaman, J. 2002. Dynamics in rhodopsin. *Chembiochem* 3:981-6.
Klein-Seetharaman, J., J. Hwa, K. Cai, C. Altenbach, W. L. Hubbell, and H. G. Khorana. 2001. Probing the Dark State Tertiary Structure in the Cytoplasmic Domain of Rhodopsin: Proximities between Amino Acids Deduced from Spontaneous Disulfide Bond Formation between Cys316 and Engineered Cysteines in Cytoplasmic Loop 1. *Biochemistry* 40:12472-12478.
Klein-Seetharaman, J. and P. J. Reeves. 2002. Solution NMR spectroscopy of [α -15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc. Natl. Acad. Sci. U. S. A.* 99:3452-7.
Klein-Seetharaman, J., P. J. Reeves, M. C. Loewen, E. V. Getmanova, J. Chung, H. Schwalbe, P. E. Wright, and H. G. Khorana. 2002. Solution NMR spectroscopy of [α -^{15}N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc. Natl. Acad. Sci. U. S. A* 99:3452-7.
Klein-Seetharaman, J., N. V. Yanamala, F. Javeed, P. J. Reeves, E. V. Getmanova, M. C. Loewen, H. Schwalbe, and H. G. Khorana. 2004. Differential dynamics in the G protein-

coupled receptor rhodopsin revealed by solution NMR. *Proc. Natl. Acad. Sci. U. S. A* 101:3409-3413.
Kono, M., H. Yu, and D. D. Oprian. 1998. Disulfide bond exchange in rhodopsin. *Biochemistry* 37:1302-5.
Kundu, S., J. S. Melton, D. C. Sorensen, and G. N. Jr. Phillips. 2002. Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models. *Biophys. J.* 83:723-732.
Lambright, D. G., J. P. Noel, H. E. Hamm, and P. B. Sigler. 1994. Structural determinants for activation of the α-subunit of a heterotrimeric G-protein. *Nature* 369:621-628.
Lambright, D. G., J. Sondek, A. Bohm, N. P. Skiba, H. E. Hamm, and P. B. Sigler. 1996. The 2.0 angstrom crystal structure of a heterotrimeric G protein
46. *Nature* 379:311-319.
Leach, A. R. 2001. Molecular Modelling: Principles and Applications. Prentice Hall.
Lemaitre, V., P. Yeagle, and A. Watts. 2005. Molecular dynamics simulations of retinal in rhodopsin: from the dark-adapted state towards lumirhodopsin. *Biochemistry* 44:12667-12680.
Liu, X., P. Garriga, and H. G. Khorana. 1996. Structure and function in rhodopsin: correct folding and misfolding in two point mutants in the intradiscal domain of rhodopsin identified in retinitis pigmentosa. *Proc. Natl. Acad. Sci. U. S. A.* 93:4554-9.
Lu, M. and J. Ma. 2005. The role of shape in determining molecular motions. *Biophys. J.* 89:2395-2401.
Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure.* 13:373-380.
Macke, J. P., C. M. Davenport, S. G. Jacobson, J. C. Hennessey, F. Gonzalez-Fernandez, B. P. Conway, J. Heckenlively, R. Palmer, I. H. Maumenee, P. Sieving, and et al. 1993. Identification of novel rhodopsin mutations responsible for retinitis pigmentosa: implications for the structure and function of rhodopsin. *Am. J. Hum. Genet.* 53:80-9.
MacKerell, A. D., Jr. 2004. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25:1584-1604.
MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Chem. Phys.* 102:3586-3616.
Marin, E. P., A. G. Krishna, V. Archambault, E. Simuni, W. Y. Fu, and T. P. Sakmar. 2001. The function of interdomain interactions in controlling nucleotide exchange rates in transducin. *J. Biol. Chem.* 276:23873-23880.
Marin, E. P., K. G. Krishna, T. A. Zvyaga, J. Isele, F. Siebert, and T. P. Sakmar. 2000. The amino terminus of the fourth cytoplasmic loop of rhodopsin modulates rhodopsin-transducin interaction. *J. Biol. Chem.* 275:1930-1936.
Massotte, D. and B. L. Kieffer. 2005. The second extracellular loop: a damper for G protein-coupled receptors? *Nat. Struct. Mol. Biol.* 12:287-288.
Meng, E. C. and H. R. Bourne. 2001. Receptor activation: what does the rhodopsin structure tell us? *Trends Pharmacol. Sci.* 22:587-93.
Menon, S. T., M. Han, and T. P. Sakmar. 2001. Rhodopsin: Structural basis of molecular physiology. *Physiol. Rev.* 81:1659-1688.

Ming, D., Y. F. Kong, M. A. Lambert, Z. Huang, and J. P. Ma. 2002. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc. Natl. Acad. Sci. U. S. A* 99:8620-8625.

Mirzadegan, T., G. Benkö, S. Filipek, and K. Palczewski. 2003. Sequence Analyses of G-Protein-Coupled Receptros: Similarities to Rhodopsin. *Biochemistry* 42:2759-67.

Müller, G. 2000. Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach. *Curr. Med. Chem.* 7:861-88.

Nakayama, T. A. and H. G. Khorana. 1990. Orientation of retinal in bovine rhodopsin determined by cross-linking using a photoactivatable analog of 11-cis-retinal. *J. Biol. Chem.* 265:15762-9.

Nakayama, T. A. and H. G. Khorana. 1991. Mapping of the amino acids in membrane-embedded helices that interact with the retinal chromophore in bovine rhodopsin. *J. Biol. Chem.* 266:4269-75.

Nikiforovich, G. V. and G. R. Marshall. 2005. Modeling flexible loops in the dark-adapted and activated states of rhodopsin, a prototypical G-protein-coupled receptor. *Biophys. J.* 89:3780-3789.

Nikiforovich, G. V. and G. R. Marshall. 2003. Three-dimensional model for Meta-II rhodopsin, an activated G-protein-coupled receptor. *Biochemistry* 42:9110-9120.

Niv, M. Y., L. Skrabanek, M. Filizola, and H. Weinstein. 2006. Modeling activated states of GPCRs: the rhodopsin template. *J. Comput. Aid. Mol. Des.* 20:437-448.

Noel, J. P., H. E. Hamm, and P. B. Sigler. 1993. The 2.2 A crystal structure of transducin-α complexed with GTPγ S. *Nature* 366:654-663.

Okada, T., Y. Fujiyoshi, M. Silow, J. Navarro, E. M. Landau, and Y. Shichida. 2002. Functional role of internal water molecules in rhodopsin revealed by X- ray crystallography. *Proc. Natl. Acad. Sci. U. S. A* 99:5982-5987.

Okada, T. and K. Palczewski. 2001. Crystal structure of rhodopsin: implications for vision and beyond. *Curr. Opin. Struct. Biol.* 11:420-6.

Okada, T., M. Sugihara, A. N. Bondar, M. Elstner, P. Entel, and V. Buss. 2004. The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. *J. Mol. Biol.* 342:571-583.

Olsson, J. E., J. W. Gordon, B. S. Pawlyk, D. Roof, A. Hayes, R. S. Molday, S. Mukai, G. S. Cowley, E. L. Berson, and T. P. Dryja. 1992. Transgenic mice with a rhodopsin mutation (Pro23His): a mouse model of autosomal dominant retinitis pigmentosa. *Neuron* 9:815-830.

Palczewski, K. 2006. G protein-coupled receptor rhodopsin. *Annu. Rev. Biochem.* 75:743-767.

Palczewski, K., T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289:739-45.

Patel, A. B., E. Crocker, M. Eilers, A. Hirshfeld, M. Sheves, and S. O. Smith. 2004. Coupling of retinal isomerization to the activation of rhodopsin. *Proc. Natl. Acad. Sci. U. S. A* 101:10048-10053.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. 1995. Amber, A Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Comm.* 91:1-41.

Peng, J. Y., N. Vaidehi, S. E. Hall, and W. A. Goddard, III. 2006. The predicted 3D structures of the human M1 muscarinic acetylcholine receptor with agonist or antagonist bound. *ChemMedChem.* 1:878-890.

Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781-1802.

Rader, A. J., G. Anderson, B. Isin, H. G. Khorana, I. Bahar, and J. Klein-Seetharaman. 2004. Identification of core amino acids stabilizing rhodopsin. *Proc. Natl. Acad. Sci. U. S. A* 101:7246-7251.

Rader, A. J., B. M. Hespenheide, L. A. Kuhn, and M. F. Thorpe. 2002. Protein unfolding: rigidity lost. *Proc. Natl. Acad. Sci. U. S. A.* 99:3540-5.

Rasmussen, S. G., A. D. Jensen, G. Liapakis, P. Ghanouni, J. A. Javitch, and U. Gether. 1999. Mutation of a highly conserved aspartic acid in the beta2 adrenergic receptor: constitutive activation, structural instability, and conformational rearrangement of transmembrane segment 6. *Mol. Pharmacol.* 56:175-84.

Ridge, K. D. and N. G. Abdulaev. 2000. Folding and assembly of rhodopsin from expressed fragments. *Methods Enzymol.* 315:59-70.

Ridge, K. D., S. Bhattacharya, T. A. Nakayama, and H. G. Khorana. 1992. Light-stable rhodopsin. II. An opsin mutant (TRP-265----Phe) and a retinal analog with a nonisomerizable 11-cis configuration form a photostable chromophore. *J. Biol. Chem.* 267:6770-5.

Rohrig, U. F., L. Guidoni, and U. Rothlisberger. 2002. Early steps of the intramolecular signal transduction in rhodopsin explored by molecular dynamics simulations. *Biochemistry* 41:10799-10809.

Saam, J., E. Tajkhorshid, S. Hayashi, and K. Schulten. 2002. Molecular dynamics investigation of primary photoinduced events in the activation of rhodopsin. *Biophys. J.* 83:3097-3112.

Sakmar, T. P. 1998. Rhodopsin: a prototypical G protein-coupled receptor. *Prog. Nucleic Acid Res. Mol. Biol.* 5:1-34.

Sakmar, T. P., R. R. Franke, and H. G. Khorana. 1991. The role of the retinylidene Schiff base counterion in rhodopsin in determining wavelength absorbance and Schiff base pKa. *Proc. Natl. Acad. Sci. U. S. A* 88:3079-83.

Sakmar, T. P., S. T. Menon, E. P. Marin, and E. S. Awad. 2002. Rhodopsin: Insights from recent structural studies. *Annu. Rev. Biophys. Biomol. Struct.* 31:443-484.

Salom, D., D. T. Lodowski, R. E. Stenkamp, T. Le, I, M. Golczak, B. Jastrzebska, T. Harris, J. A. Ballesteros, and K. Palczewski. 2006. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U. S. A* 103:16123-16128.

Samama, P., S. Cotecchia, T. Costa, and R. J. Lefkowitz. 1993. A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model. *J Biol Chem* 268:4625-36.

Sanchez, R., U. Pieper, F. Melo, N. Eswar, M. A. Marti-Renom, M. S. Madhusudhan, N. Mirkovic, and A. Sali. 2000. Protein structure modeling for structural genomics. *Nat. Struct. Biol.* 7 Suppl:986-990.

Scheer, A., F. Fanelli, T. Costa, P. G. De Benedetti, and S. Cotecchia. 1996. Constitutively active mutants of the $\alpha_1\beta$-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *Embo J.* 15:3566-78.

Schlick, T. 2002. Molecular modeling and simulation: an interdisciplinary guide. Springer-Verlag, New York.

Schlitter, J., M. Engels, and P. Kruger. 1994. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* 12:84-89.

Scott, W. R. P., P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren. 1999. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* 103:3596-3607.

Shi, L., G. Liapakis, R. Xu, F. Guarnieri, J. A. Ballesteros, and J. A. Javitch. 2002. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J. Biol. Chem.* 277:40989-96.

Shrivastava, I. H. and I. Bahar. 2006. Common mechanism of pore opening shared by five different potassium channels. *Biophys. J.* 90:3929-3940.

Suel, G. M., S. W. Lockless, M. A. Wall, and R. Ranganathan. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10:59-69.

Sung, C. H., C. M. Davenport, J. C. Hennessey, I. H. Maumenee, S. G. Jacobson, J. R. Heckenlively, R. Nowakowski, G. Fishman, P. Gouras, and J. Nathans. 1991. Rhodopsin mutations in autosomal dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci. U. S. A.* 88:6481-5.

Tama, F. and Y. H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1-6.

Teller, D. C., T. Okada, C. A. Behnke, K. Palczewski, and R. E. Stenkamp. 2001. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry* 40:7761-72.

Tellez-Sanz, R., E. Cesareo, M. Nuccetelli, A. M. Aguilera, C. Baron, L. J. Parker, J. J. Adams, C. J. Morton, B. M. Lo, M. W. Parker, and L. Garcia-Fuentes. 2006. Calorimetric and structural studies of the nitric oxide carrier S-nitrosoglutathione bound to human glutathione transferase P1-1. *Protein Sci.* 15:1093-1105.

Tirion, M. M. 1996. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* 77:1905-1908.

Tobi, D. and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. U. S. A* 102:18908-18913.

Unger, V. M. and G. F. X. Schertler. 1995. Low-resolution structure of bovine rhodopsin determined by electron cryomicroscopy. *Biophys. J.* 68:1776-1786.

Vaidehi, N., W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos, and W. A. Goddard, III. 2002. Prediction of structure and function of G protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A* 99:12622-12627.

Vaidehi, N., S. Schlyer, R. J. Trabanino, W. B. Floriano, R. Abrol, S. Sharma, M. Kochanny, S. Koovakat, L. Dunning, M. Liang, J. M. Fox, F. L. de Mendonca, J. E. Pease, W. A. Goddard, III, and R. Horuk. 2006. Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. *J. Biol. Chem.* 281:27613-27620.

Van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 26:1701-1718.

van Gunsteren, W. F., D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hunenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, d. van, V, and H. B. Yu. 2006. Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed. Engl.* 45:4064-4092.

Vandrunen, R., D. Vanderspoel, and H. J. C. Berendsen. 1995. Gromacs - A Software Package and A Parallel Computer for Molecular-Dynamics. *Abstracts of Papers. Am. Chem. Soc.* 209:49-COMP.

Visiers, I., J. A. Ballesteros, and H. Weinstein. 2002. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Methods Enzymol.* 343:329-71.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics* 8:52-6.

Wald, G. 1953. The biochemistry of vision. *Annu. Rev. Biochem.* 22:497-526.

Wald, G. 1952. Retinal mechanism of vision. *Am. J. Ophthalmol.* 35:1505-1507.

Wang, M., T. T. Lam, M. O. Tso, and M. I. Naash. 1997. Expression of a mutant opsin gene increases the susceptibility of the retina to light damage. *Vis. Neurosci.* 14:55-62.

Williams, M. A., J. M. Goodfellow, and J. M. Thornton. 1994. Buried waters and internal cavities in monomeric proteins. *Protein Sci.* 3:1224-35.

Xu, C., D. Tobi, and I. Bahar. 2003. Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T<-> R2 Transition. J. Mol. Biol. 333:153-168.

Yang, L. W. and I. Bahar. 2005. Coupling between Catalytic Site and Collective Dynamics: A requirement for Mechanochemical Activity of Enzymes. Structure.

Zhang, Z., Y. Shi, and H. Liu. 2003. Molecular dynamics simulations of peptides and proteins with amplified collective motions. Biophys. J. 84:3583-3593.